SAS/ETS® 14.2 User's Guide
The QLIM Procedure

This document is an individual chapter from *SAS/ETS® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/ETS® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/ETS® 14.2 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

Chapter 28
The QLIM Procedure

Contents

Overview: QLIM Procedure

The QLIM (qualitative and limited dependent variable model) procedure analyzes univariate and multivariate limited dependent variable models in which dependent variables take discrete values or in which dependent variables are observed only in a limited range of values. These models include logit, probit, tobit, selection, and multivariate models. The multivariate model can contain discrete choice and limited endogenous variables in addition to continuous endogenous variables.

The QLIM procedure supports the following models:

- linear regression model with heteroscedasticity
- Box-Cox regression with heteroscedasticity
- probit with heteroscedasticity
- logit with heteroscedasticity
- tobit (censored and truncated) with heteroscedasticity
- bivariate probit
- bivariate tobit
- sample selection and switching regression models

- multivariate limited dependent variables
- stochastic frontier production and cost models

In the linear regression models with heteroscedasticity, the assumption that error variance is constant across observations is relaxed. The QLIM procedure allows for a number of different linear and nonlinear variance specifications. Another way to make the linear model more appropriate to fit the data and reduce skewness is to apply Box-Cox transformation. If the nature of the data is such that the dependent variable is discrete and it takes only two possible values, ordinary least squares (OLS) estimates are inconsistent. The QLIM procedure offers probit and logit models to overcome these estimation problems. Assumptions about the error variance can also be relaxed in order to estimate probit or logit with heteroscedasticity.

The QLIM procedure also offers a class of models in which the dependent variable is censored or truncated from below or above or both. When a continuous dependent variable is observed only within a certain range and values outside this range are not available, the QLIM procedure offers a class of models that adjust for truncation. In some cases, the dependent variable is continuous only in a certain range and all values outside this range are reported as being on its boundary. For example, if it is not possible to observe negative values, the value of the dependent variable is reported as equal to 0. Because the data are censored, OLS results are inconsistent, and it cannot be guaranteed that the predicted values from the model fall in the appropriate region.

Most of the models in the QLIM procedure can be extended to accommodate bivariate and multivariate scenarios. The assumption that one variable is observed only if another variable takes on certain values lead to the introduction of sample selection models. If the dependent variables are mutually exclusive and observed only for certain ranges of the selection variable, the sample selection can be extended to include cases of switching regression. Stochastic frontier production and cost models allow for random shocks of the production or cost. They include a systematic positive component in the error term that adjusts for technological or cost inefficiency.

The QLIM procedure can use the maximum likelihood method and the Bayesian method for both univariate and multivariate models. Initial starting values for the nonlinear optimizations are typically calculated by OLS.

Getting Started: QLIM Procedure

The QLIM procedure is similar in use to the other regression or simultaneous equations model procedures in the SAS System. For example, the following statements are used to estimate a binary choice model by using the probit probability function:

```
proc qlim data=a;
   model y = x1;
   endogenous y ~ discrete;
run;
```

The response variable, y, is numeric and has discrete values. PROC QLIM enables the user to specify the type of endogenous variables in the ENDOGENOUS statement. The binary probit model can be also specified as follows:

```
model y = x1 / discrete;
```

When multiple endogenous variables are specified in the QLIM procedure, these equations are estimated as a system. Multiple endogenous variables can be specified with one MODEL statement in the QLIM procedure when these models have the same exogenous variables:

```
model y1 y2 = x1 x2 / discrete;
```

The preceding specification is equivalent to the following statements:

```
proc qlim data=a;
   model y1 = x1 x2;
   model y2 = x1 x2;
   endogenous y1 y2 ~ discrete;
run;
```

Some equations in multivariate models can be continuous while other equations can be discrete. A bivariate model with a discrete and a continuous equation is specified as follows:

```
proc qlim data=a;
   model y1 = x1 x2;
   model y2 = x3 x4;
   endogenous y1 ~ discrete;
run;
```

The standard tobit model is estimated by specifying the endogenous variable to be truncated or censored. The limits of the dependent variable can be specified with the CENSORED or TRUNCATED option in the ENDOGENOUS or MODEL statement when the data are limited by specific values or variables. For example, the two-limit censored model requires two variables that contain the lower (bottom) and upper (top) bound:

```
proc qlim data=a;
   model y = x1 x2 x3;
   endogenous y ~ censored(lb=bottom ub=top);
run;
```

The bounds can be numbers if they are fixed for all observations in the data set. For example, the standard tobit model can be specified as follows:

```
proc qlim data=a;
   model y = x1 x2 x3;
   endogenous y ~ censored(lb=0);
run;
```

Introductory Example: Binary Probit and Logit Models

The following example illustrates the use of PROC QLIM. The data were originally published by Mroz (1987) and downloaded from Wooldridge (2002). This data set is based on a sample of 753 married white women. The dependent variable is a discrete variable of labor force participation (inlf). Explanatory variables are the number of children ages 5 or younger (kidslt6), the number of children ages 6 to 18 (kidsge6), the woman's age (age), the woman's years of schooling (educ), wife's labor experience (exper), square of

experience (expersq), and the family income excluding the wife's wage (nwifeinc). The program (with data values omitted) is as follows:

```
/*-- Binary Probit --*/
proc qlim data=mroz plots=predicted;
   model inlf = nwifeinc educ exper expersq
                age kidslt6 kidsge6  / discrete;
run;
```

Results of this analysis are shown in the following four figures. In the first table, shown in Figure 28.1, PROC QLIM provides frequency information about each choice. In this example, 428 women participate in the labor force (inlf = 1).

Figure 28.1 Choice Frequency Summary

Binary Data

The QLIM Procedure

Discrete Response Profile of inlf		
Index	Value	Total Frequency
1	0	325
2	1	428

The second table is the estimation summary table shown in Figure 28.2. Included are the number of dependent variables, names of dependent variables, the number of observations, the log-likelihood function value, the maximum absolute gradient, the number of iterations, AIC, and Schwarz criterion.

Figure 28.2 Fit Summary Table of Binary Probit

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	inlf
Number of Observations	753
Log Likelihood	-401.30219
Maximum Absolute Gradient	0.0000669
Number of Iterations	15
Optimization Method	Quasi-Newton
AIC	818.60439
Schwarz Criterion	855.59691

Goodness-of-fit measures are displayed in Figure 28.3. All measures except McKelvey-Zavoina's definition are based on the log-likelihood function value. The likelihood ratio test statistic has chi-square distribution conditional on the null hypothesis that all slope coefficients are zero. In this example, the likelihood ratio statistic is used to test the hypothesis that kidslt6 =kidge6 =age =educ =exper =expersq =nwifeinc = 0.

Figure 28.3 Goodness of Fit

Goodness-of-Fit Measures

Measure	Value	Formula
Likelihood Ratio (R)	227.14	2 * (LogL - LogL0)
Upper Bound of R (U)	1029.7	- 2 * LogL0
Aldrich-Nelson	0.2317	R / (R+N)
Cragg-Uhler 1	0.2604	1 - exp(-R/N)
Cragg-Uhler 2	0.3494	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.2888	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.2693	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2206	R / U
Veall-Zimmermann	0.4012	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.4025	

N = # of observations, K = # of regressors

The parameter estimates and standard errors are shown in Figure 28.4.

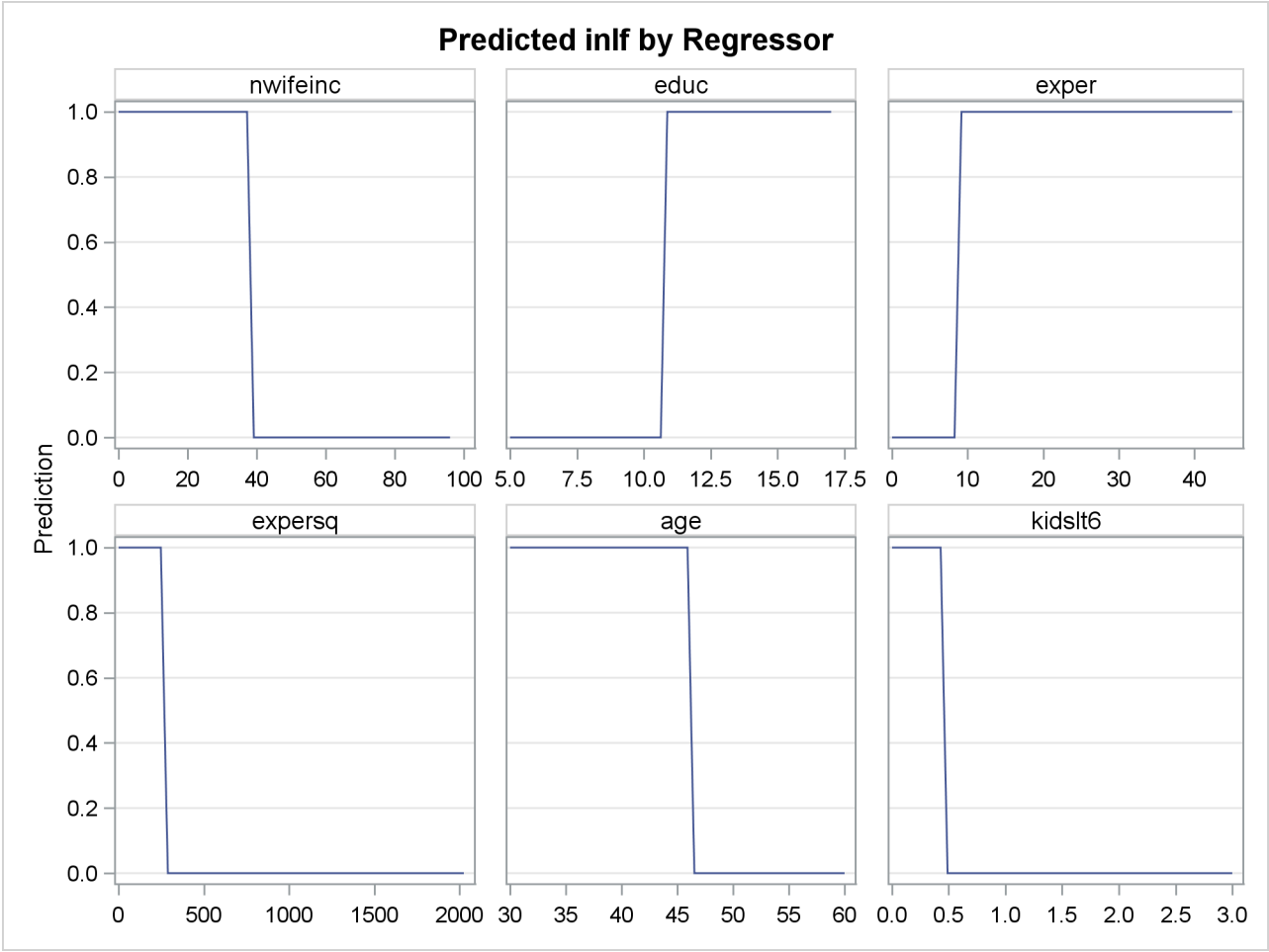
Figure 28.4 Parameter Estimates of Binary Probit

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.270077	0.508590	0.53	0.5954
nwifeinc	1	-0.012024	0.004840	-2.48	0.0130
educ	1	0.130905	0.025255	5.18	<.0001
exper	1	0.123348	0.018720	6.59	<.0001
expersq	1	-0.001887	0.000600	-3.14	0.0017
age	1	-0.052853	0.008477	-6.24	<.0001
kidslt6	1	-0.868329	0.118519	-7.33	<.0001
kidsge6	1	0.036005	0.043477	0.83	0.4076

Finally, the QLIM procedure profiles the predicted outcome with respect to the regressors. For example, Output 28.5 shows the predicted values profiled with respect to nwifeinc, educ, exper, expersq, age, and kidslt6.

Figure 28.5 Predictions by Regressors: nwifeinc, educ, exper, expersq, age, and kidslt6



When the error term has a logistic distribution, the binary logit model is estimated. To specify a logistic distribution, add the D=LOGIT option as follows:

```
/*-- Binary Logit --*/
proc qlim data=mroz;
   model inlf = nwifeinc educ exper expersq
                age kidslt6 kidsge6  / discrete(d=logit);
run;
```

The estimated parameters are shown in Figure 28.6.

Figure 28.6 Parameter Estimates of Binary Logit

Binary Data

The QLIM Procedure

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.425452	0.860365	0.49	0.6210
nwifeinc	1	-0.021345	0.008421	-2.53	0.0113
educ	1	0.221170	0.043441	5.09	<.0001
exper	1	0.205870	0.032070	6.42	<.0001
expersq	1	-0.003154	0.001017	-3.10	0.0019
age	1	-0.088024	0.014572	-6.04	<.0001
kidslt6	1	-1.443354	0.203575	-7.09	<.0001
kidsge6	1	0.060112	0.074791	0.80	0.4215

The heteroscedastic logit model can be estimated using the HETERO statement. If the variance of the logit model is a function of the family income level excluding wife's income (nwifeinc), the variance can be specified as

$$\mathrm{Var}(\epsilon_i) = \sigma^2 \exp(\gamma * \mathrm{nwifeinc}_i)$$

where σ^2 is normalized to 1 because the dependent variable is discrete. The following SAS statements estimate the heteroscedastic logit model:

```
/*-- Binary Logit with Heteroscedasticity --*/
proc qlim data=mroz;
  model inlf = nwifeinc educ exper expersq
               age kidslt6 kidsge6 / discrete(d=logit);
  hetero inlf ~ nwifeinc / noconst;
run;
```

The parameter estimate, γ, of the heteroscedasticity variable is listed as _H.nwifeinc; see Figure 28.7.

Figure 28.7 Parameter Estimates of Binary Logit with Heteroscedasticity

Binary Data

The QLIM Procedure

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.510445	0.983538	0.52	0.6038
nwifeinc	1	-0.026778	0.012108	-2.21	0.0270
educ	1	0.255547	0.061728	4.14	<.0001
exper	1	0.234105	0.046639	5.02	<.0001
expersq	1	-0.003613	0.001236	-2.92	0.0035
age	1	-0.100878	0.021491	-4.69	<.0001
kidslt6	1	-1.645206	0.311296	-5.29	<.0001
kidsge6	1	0.066941	0.085633	0.78	0.4344
_H.nwifeinc	1	0.013280	0.013606	0.98	0.3291

Syntax: QLIM Procedure

The following statements are available in the QLIM procedure:

PROC QLIM < *options* > ;
BAYES < *options* > ;
BOUNDS *bound1* < , *bound2* . . . > ;
BY *variables* ;
CLASS *variables* ;
FREQ *variable* ;
ENDOGENOUS *variables* ~ *options* ;
HETERO *dependent-variables* ~ *exogenous-variables* / *options* ;
INIT *initvalue1* < , *initvalue2* . . . > ;
MODEL *dependent-variable* = *regressors* / *options* ;
NLOPTIONS < *options* > ;
OUTPUT < **OUT**=*SAS-data-set* > < *output-options* > ;
PRIOR *parameter-list* ~ *distribution* ;
RANDOM *regressors* < / *options* > ;
RESTRICT *restriction1* < , *restriction2* . . . > ;
TEST *options* ;
WEIGHT *variable* < / *options* > ;

At least one MODEL statement is required. If more than one MODEL statement is used, the QLIM procedure estimates a system of models. If a FREQ or WEIGHT statement is specified more than once, the variable specified in the first instance is used. Main effects and higher-order terms can be specified in the MODEL statement, as in the GLM procedure and PROBIT procedure in SAS/STAT. If a CLASS statement is used, it must precede the MODEL statement.

Functional Summary

Table 28.1 summarizes the statements and options used with the QLIM procedure.

Table 28.1 Functional Summary

Description	Statement	Option
Data Set Options		
Specifies the input data set	PROC QLIM	DATA=
Writes parameter estimates to an output data set	PROC QLIM	OUTEST=
Writes predictions to an output data set	OUTPUT	OUT=
Declaring the Role of Variables		
Specifies BY-group processing	BY	
Specifies classification variables	CLASS	
Specifies a frequency variable	FREQ	
Specifies a weight variable	WEIGHT	NONORMALIZE
Printing Control Options		
Requests all printing options	PROC QLIM	PRINTALL
Prints correlation matrix of the estimates	PROC QLIM	CORRB
Prints covariance matrix of the estimates	PROC QLIM	COVB
Prints a summary iteration listing	PROC QLIM	ITPRINT
Suppresses the normal printed output	PROC QLIM	NOPRINT
Plotting Options		
Displays plots	PROC QLIM	PLOTS=
Options to Control the Optimization Process		
Specifies the optimization method	PROC QLIM	METHOD=
Specifies the optimization options	NLOPTIONS	See Chapter 6, "Nonlinear Optimization Methods."
Sets initial values for parameters	INIT	
Specifies upper and lower bounds for the parameter estimates	BOUNDS	
Specifies linear restrictions on the parameter estimates	RESTRICT	
Model Estimation Options		
Specifies options specific to Box-Cox transformation	MODEL	BOXCOX()
Suppresses the intercept parameter	MODEL	NOINT
Specifies variable selection	MODEL	SELECTVAR=( )
Specifies the type of random number generators	MODEL	RANDNUM=
Specifies that initial values are generated using random numbers	MODEL	RANDOMINIT

Table 28.1 *continued*

Description	Statement	Option
Specifies a seed for pseudorandom number generation	PROC QLIM	SEED=
Specifies the number of draws for Monte Carlo integration	PROC QLIM	NDRAW=
Specifies the method to calculate parameter covariance	PROC QLIM	COVEST=
Requests estimation by Heckman's two-step method	PROC QLIM	HECKIT
Options for the Estimation of Random-Parameters Models		
Specifies the ID variable for the parameter heterogeneity	RANDOM	SUBJECT=
Requests the MC simulation method of integration	RANDOM	METHOD=SIMULATION()
Requests the Halton sequence method of integration	RANDOM	METHOD=HALTON()
Requests the Gauss-Hermite quadrature method of integration	RANDOM	METHOD=HERMITE()
Requests that random parameters be uncorrelated	RANDOM	NOCORRELATION
Bayesian MCMC Options		
Controls the aggregation of multiple posterior chains	BAYES	AGGREGATION=
Automates the initialization of the MCMC algorithm	BAYES	AUTOMCMC()
Specifies the initial values of the MCMC	INIT	
Evaluates the marginal likelihood	BAYES	MARGINLIKE
Specifies the maximum number of tuning phases	BAYES	MAXTUNE=
Specifies the minimum number of tuning phases	BAYES	MINTUNE=
Specifies the number of burn-in iterations	BAYES	NBI=
Specifies the number of iterations during the sampling phase	BAYES	NMC=
Specifies the number of samples for the prior predictive analysis	BAYES	NMCPRIOR=
Specifies the number of threads to use during the sampling phase	BAYES	NTRDS=
Specifies the number of iterations during the tuning phase	BAYES	NTU=
Controls options for constructing the initial proposal covariance matrix	BAYES	PROPCOV=
Specifies the sampling scheme	BAYES	SAMPLING=
Specifies the random number generator seed	BAYES	SEED=
Prints the time required for the MCMC sampling	BAYES	SIMTIME
Controls the thinning of the Markov chain	BAYES	THIN=
Bayesian Summary Statistics and Convergence Diagnostics		
Displays convergence diagnostics	BAYES	DIAGNOSTICS=

Table 28.1 *continued*

Description	Statement	Option
Displays summary statistics of the posterior samples	BAYES	STATISTICS=
Bayesian Prior and Posterior Samples		
Specifies a SAS data set for the posterior samples	BAYES	OUTPOST=
Specifies a SAS data set for the prior samples	BAYES	OUTPRIOR=
Bayesian Analysis		
Specifies normal prior distribution	PRIOR	NORMAL(MEAN=, VAR=)
Specifies gamma prior distribution	PRIOR	GAMMA(SHAPE=, SCALE=)
Specifies square root gamma prior distribution	PRIOR	SQGAMMA(SHAPE=, SCALE=)
Specifies inverse gamma prior distribution	PRIOR	IGAMMA(SHAPE=, SCALE=)
Specifies square root inverse gamma prior distribution	PRIOR	SQIGAMMA(SHAPE=, SCALE=)
Specifies uniform prior distribution	PRIOR	UNIFORM(MIN=, MAX=)
Specifies beta prior distribution	PRIOR	BETA(SHAPE1=, SHAPE2=, MIN=, MAX=)
Specifies *t* prior distribution	PRIOR	T(LOCATION=, DF=)
Endogenous Variable Options		
Specifies discrete variable	ENDOGENOUS	DISCRETE()
Specifies censored variable	ENDOGENOUS	CENSORED()
Specifies truncated variable	ENDOGENOUS	TRUNCATED()
Specifies variable selection condition	ENDOGENOUS	SELECT()
Specifies stochastic frontier variable	ENDOGENOUS	FRONTIER()
Endogeneity and Overidentification Test Options		
Requests the variable addition test for endogeneity	ENDOGENOUS	ENDOTEST()
Requests the overidentification test	ENDOGENOUS	OVERID()
Heteroscedasticity Model Options		
Specifies the function for heteroscedasticity models	HETERO	LINK=
Squares the function for heteroscedasticity models	HETERO	SQUARE
Specifies no constant for heteroscedasticity models	HETERO	NOCONST
Output Control Options		
Outputs predicted values	OUTPUT	PREDICTED
Outputs structured part	OUTPUT	XBETA
Outputs residuals	OUTPUT	RESIDUAL
Outputs error standard deviation	OUTPUT	ERRSTD
Outputs marginal effects	OUTPUT	MARGINAL
Outputs probability for the current response	OUTPUT	PROB
Outputs probability for all responses	OUTPUT	PROBALL

Table 28.1 *continued*

Description	Statement	Option
Outputs expected value	OUTPUT	EXPECTED
Outputs conditional expected value	OUTPUT	CONDITIONAL
Outputs inverse Mills ratio	OUTPUT	MILLS
Outputs technical efficiency measures	OUTPUT	TE1
	OUTPUT	TE2
Includes covariances in the OUTEST= data set	PROC QLIM	COVOUT
Includes correlations in the OUTEST= data set	PROC QLIM	CORROUT
Test Request Options		
Requests Wald, Lagrange multiplier, and likelihood ratio tests	TEST	ALL
Requests the Wald test	TEST	WALD
Requests the Lagrange multiplier test	TEST	LM
Requests the likelihood ratio test	TEST	LR

PROC QLIM Statement

PROC QLIM < *options* > **;**

You can specify the following *options* in the PROC QLIM statement.

Data Set Options

DATA=*SAS-data-set*
specifies the input SAS data set. If this option is not specified, PROC QLIM uses the most recently created SAS data set.

Output Data Set Options

OUTEST=*SAS-data-set*
writes the parameter estimates to the specified *SAS-data-set*.

COVOUT
writes the covariance matrix for the parameter estimates to the OUTEST= data set. This option is valid only if the OUTEST= option is specified.

CORROUT
writes the correlation matrix for the parameter estimates to the OUTEST= data set. This option is valid only if the OUTEST= option is specified.

Printing Options

NOPRINT
suppresses the normal printed output but does not suppress error listings. If you specify the NOPRINT option, then any other print option is turned off.

PRINTALL
turns on all the printing-control options. The options set by PRINTALL are COVB and CORRB.

CORRB
prints the correlation matrix of the parameter estimates.

COVB
prints the covariance matrix of the parameter estimates.

ITPRINT
prints the initial parameter estimates, convergence criteria, and all constraints of the optimization. At each iteration, objective function value, step size, maximum gradient, and slope of search direction are printed as well.

Model Estimation Options

COVEST=OP | HESSIAN | QML
specifies the method for calculating the covariance matrix of parameter estimates. You can specify the following *covariance-options*:

OP	calculates the covariance from the outer product matrix.
HESSIAN	calculates the covariance from the inverse Hessian matrix.
QML	calculates the covariance from the outer product and Hessian matrices (the quasi-maximum likelihood estimates).

By default, COVEST=HESSIAN.

HECKIT < (*heckit-options*) >
uses Heckman's two-step estimation method to estimate the selection model. You must specify exactly two MODEL statements when you use the HECKIT option. One of the models must be a binary probit model; therefore, you must specify the DISCRETE option in the MODEL or ENDOGENOUS statement. You base the selection on the binary probit model for the second model; therefore, you must specify the SELECT option for this model.

You can specify one or both of the following *heckit-options*:

SECONDSTAGE=OLS | ML
specifies the estimation method of the second stage of Heckman's two-step method. You can specify the following values:

OLS requests the ordinary least squares method for the second stage. If you specify SECONDSTAGE=OLS, then the model of interest—that is, the model that uses the SELECT option—must be linear and contain a continuous dependent variable. Therefore, you cannot specify the DISCRETE, CENSORED, or TRUNCATED option

along with the SELECT option for the model of interest. When you specify SECONDSTAGE=OLS, you cannot test or restrict the parameters of the model of interest. However, you can test or restrict the parameters of the selection model—that is, the model that defines the selection rule.

ML requests that PROC QLIM use the maximum likelihood method in the second stage, as it does in the first stage. When you specify SECONDSTAGE=ML, the model of interest can be nonlinear. Moreover, you can also use the TEST or RESTRICT statement to test or restrict the parameters of the model of interest.

By default, SECONDSTAGE=OLS.

UNCORRECTED
requests the conventional OLS standard errors when the second-stage estimation method is the ordinary least squares method. If you do not specify the UNCORRECTED option, PROC QLIM reports the corrected OLS standard errors. For more information about the corrected standard errors, see the section "Heckman's Two-Step Selection Method" on page 1998.

If you specify both the UNCORRECTED and SECONDSTAGE=ML options, PROC QLIM ignores the UNCORRECTED option, because the UNCORRECTED option is related to the OLS standard errors.

NDRAW=*value*
specifies the number of draws for Monte Carlo integration.

SEED=*value*
specifies a seed for pseudorandom number generation in Monte Carlo integration.

Optimization Process Control Options

PROC QLIM uses the nonlinear optimization (NLO) subsystem to perform nonlinear optimization tasks. You can use any of the NLO options in the NLOPTIONS statement. For more information, see Chapter 6, "Nonlinear Optimization Methods."

METHOD=*value*
specifies the optimization method. If this option is specified, it overwrites the TECH= option in the NLOPTIONS statement. You can specify the following *values*:

CONGRA performs a conjugate-gradient optimization.

DBLDOG performs a version of double-dogleg optimization.

NEWRAP performs a Newton-Raphson optimization, combining a line-search algorithm with ridging.

NMSIMP performs a Nelder-Mead simplex optimization.

NONE specifies that no optimization be performed beyond using the ordinary least squares method to compute the parameter estimates.

NRRIDG performs a Newton-Raphson optimization with ridging.

QUANEW performs a quasi-Newton optimization.

TRUREG performs a trust region optimization.

By default, METHOD=QUANEW.

Plotting Options

PLOTS< (*global-plot-options*) > = *plot-request* | *(plot-requests)*
controls the display of plots. By default, the plots are displayed in panels unless the UNPACK *global-plot-option* is specified. When you specify only one *plot-request*, you can omit the parentheses around the *plot-request*.

Global Plot Options

You can specify the following *global-plot-options*:

ONLY
displays only the requested plot.

PRIOR
displays the prior predictive graph that is associated with the requested posterior predictive plot BAYESPRED. This option is available only for Bayesian analysis.

UNPACKPANEL

UNPACK
specifies that all paneled plots be unpacked, meaning that each plot in a panel is displayed separately.

Plot Requests

You can specify the following *plot-requests*:

ALL
specifies all types of available plots.

AUTOCORR< (LAGS=*n*) >
displays the autocorrelation function plots for the parameters. This *plot-request* is available only for Bayesian analysis. The optional LAGS= suboption specifies the number (up to lag *n*) of autocorrelations to be plotted in the AUTOCORR plot. If this suboption is not specified, autocorrelations are plotted up to lag 50.

BAYESDIAG
displays the TRACE, AUTOCORR, and DENSITY plots. This *plot-request* is available only for Bayesian analysis.

BAYESPRED
displays the predictive analysis. The predictive analysis takes into account the variability of the error term, whereas the PREDICTED *plot-request* does not. The BAYESPRED *plot-request* is available only for Bayesian analysis.

BAYESSUM
displays the posterior distribution, the prior distribution, and the maximum likelihood estimates. This *plot-request* is available only for Bayesian analysis.

CONDITIONAL
displays the conditional expected values for continuous endogenous variables. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

DENSITY< (FRINGE) >
displays the kernel density plots for the parameters. This *plot-request* is available only for Bayesian analysis. If you specify the FRINGE suboption, a fringe plot is created on the X axis of the kernel density plot. This *plot-request* is available only for Bayesian analysis.

ERRSTD
displays the error standard deviation versus observed regressors when you also specify a HETERO statement. This *plot-request* is not available for Bayesian analysis.

EXPECTED
displays the expected values for continuous endogenous variables. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

MARGINAL
displays the marginal effects. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

MILLS
displays the inverse Mills ratio. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

NONE
suppresses all diagnostic plots.

PREDICTED
displays the model predicted values. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

PROB
displays the predicted response probability. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

PROBALL
displays the predicted probabilities for each level of the response. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

PROFLIK
displays the profiled log likelihood. Each profiled graph is obtained by setting all the parameters to their maximum likelihood estimate except for the profiling parameter. The profiling parameter takes values on a predefined grid that is determined by the maximum likelihood estimate of the corresponding standard deviation. When a restricted optimization is requested, the profiled log likelihood plots depict the behavior of the profiled log likelihood around the restricted MLE without imposing the actual restrictions.

RESIDUAL
displays the residuals versus observed regressors. This *plot-request* is not available for Bayesian analysis.

TE1
displays the technical efficiency for the stochastic frontier model as suggested by Battese and Coelli (1988). Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

TE2
displays the technical efficiency for the stochastic frontier model as suggested by Jondrow et al. (1982). Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This *plot-request* is not available for Bayesian analysis.

TRACE< (SMOOTH) >
displays the trace plots for the parameters. This *plot-request* is available only for Bayesian analysis. The SMOOTH suboption displays a fitted penalized B-spline curve for each TRACE plot.

XBETA
displays the structural part on the right-hand side of the model. Each contributing regressor is set equal to its mean, except for the parameter that is reported on the X axis. This is not available for Bayesian analysis.

BAYES Statement

BAYES < *options* > **;**

The BAYES statement controls the Metropolis sampling scheme that is used to obtain samples from the posterior distribution of the underlying model and data.

AGGREGATION=WEIGHTED | UNWEIGHTED (Experimental)
specifies how multiple posterior samples should be aggregated. You can specify the following values:

WEIGHTED implements a weighted resampling scheme for the aggregation of multiple posterior chains. You can use this option when the posterior distribution is characterized by several very distinct posterior modes.

UNWEIGHTED aggregates multiple posterior chains without any adjustment. You can use this option when the posterior distribution is characterized by one or few relatively close posterior modes.

By default, AGGREGATION=UNWEIGHTED. For more information, see the section "Aggregation of Multiple Chains" on page 2017.

AUTOMCMC< =(*automcmc-options***) >** (Experimental)
specifies an algorithm for the auto-initialization of the MCMC sampling algorithm. For more information, see the section "Automated Initialization of MCMC" on page 2018.

ACCURACY=(*accuracy-options***)**
customizes the behavior of the AUTOMCMC algorithm when you are searching for an accurate representation of the posterior distribution. You can specify the following *accuracy-options*:

ATTEMPTS=*number*
specifies the maximum number of attempts that is required in order to obtain accurate samples from the posterior distribution. By default, ATTEMPTS=10.

TARGETESS=*number*
requests that the accuracy search be based on the effective sample size (ESS) analysis. If you specify this option, you must also specify the minimum *number* of effective samples.

TARGETSTATS<=(*targetstats-option***)>**
requests that the accuracy search be based on the analysis of the posterior mean and a posterior quantile of interest. You can customize the behavior of the analysis of the posterior mean by adjusting HEIDELBERGER sub-options. You can customize the behavior of the analysis of the posterior quantile by adjusting the RAFTERY sub-options. If you specify TARGETSTATS, you can also specify how the Raftery-Lewis test should be interpreted by using the following *targetstats-option*:

RLLIMITS=(LB=*number* **UB=***number***)**
specifies a region where the search for the optimal sample size depends directly on the Raftery-Lewis test. By default, RLLIMITS (LB=10000 UB=300000).

TOL=*value*
specifies the proportion of parameters that are required to be accurate. By default, TOL=0.95.

MAXNMC=*number*
specifies the maximum number of posterior samples that the AUTOMCMC option allows. By default, MAXNMC=700000.

RANDINIT< =(*randinit-options***) >**
specifies random starting points for the MCMC algorithm. The starting points can be sampled around the maximum likelihood estimate and around the prior mean. You can specify the following *randinit-options*:

MULTIPLIER=(*value***)**
specifies the radius of the area where the starting points are sampled. For the starting points that are sampled around the maximum likelihood estimate, the radius equals the standard deviation of the maximum likelihood estimate multiplied by the multiplier value. For the starting points that are sampled around the prior mean, the radius equals the standard deviation of the prior distribution multiplied by the multiplier value. By default, MULTIPLIER=2.

PROPORTION=(*value***)**
specifies the proportion of starting points that are sampled around the maximum likelihood estimate and around the prior mean. By default, PROPORTION=0, which implies that all the initial points are sampled around the maximum likelihood estimate. If you use choose to sample starting points around the prior mean, the convergence of the MCMC algorithm could be very slow.

STATIONARITY=(*stationarity-options***)**
customizes the behavior of the AUTOMCMC algorithm when you are trying to sample from the posterior distribution. You can specify the following *stationarity-options*:

ATTEMPTS=*number*
specifies the maximum number of attempts that are required in order to obtain stationary samples from the posterior distribution. By default, ATTEMPTS=10.

TOL=*value*
specifies the proportion of parameter whose samples must to be stationary. By default, TOL=0.95.

DIAGNOSTICS=ALL | NONE | *(keyword-list)*
DIAG=ALL | NONE | *(keyword-list)*
controls which diagnostics are produced. All the following diagnostics are produced with DIAGNOSTICS=ALL. If you do not want any of these diagnostics, specify DIAGNOSTICS=NONE. If you want some but not all of the diagnostics, or if you want to change certain settings of these diagnostics, specify a subset of the following keywords. By default, DIAGNOSTICS=NONE.

AUTOCORR <(LAGS=*numeric-list***)>**
computes the autocorrelations at lags that are specified in the *numeric-list*. Elements in the *numeric-list* are truncated to integers, and repeated values are removed. If the LAGS= option is not specified, autocorrelations of lags 1, 5, 10, and are computed.

AUTOMCMCSUM
produces a summary table for the AUTOMCMC (automatic MCMC) sampling tool is used.

ESS
computes Carlin's estimate of the effective sample size, the correlation time, and the efficiency of the chain for each parameter.

GEWEKE <(*geweke-options***)>**
computes the Geweke spectral density diagnostics, which are essentially a two-sample t test between the first f_1 portion and the last f_2 portion of the chain. The default is $f_1 = 0.1$ and $f_2 = 0.5$, but you can choose other fractions by using the following *geweke-options*:

FRAC1=*value*
specifies the fraction f_1 for the first window.

FRAC2=*value*
specifies the fraction f_2 for the second window.

HEIDELBERGER <(*heidel-options***)>**
computes the Heidelberger and Welch diagnostic for each variable, which consists of a stationarity test of the null hypothesis that the sample values form a stationary process. If the stationarity test is not rejected, a halfwidth test is then carried out. Optionally, you can specify one or more of the following *heidel-options*:

SALPHA=*value*
specifies the α level ($0 < \alpha < 1$) for the stationarity test.

HALPHA=*value*
specifies the α level ($0 < \alpha < 1$) for the halfwidth test.

EPS=*value*
specifies a positive number ϵ such that if the halfwidth is less than ϵ times the sample mean of the retained iterates, the halfwidth test is passed.

MCSE

MCERROR
computes the Monte Carlo standard error for each parameter. The Monte Carlo standard error, which measures the simulation accuracy, is the standard error of the posterior mean estimate and is calculated as the posterior standard deviation divided by the square root of the effective sample size.

RAFTERY<(*raftery-options***)>**
computes the Raftery and Lewis diagnostics, which evaluate the accuracy of the estimated quantile ($\hat{\theta}_Q$ for a given $Q \in (0, 1)$) of a chain. $\hat{\theta}_Q$ can achieve any degree of accuracy when the chain is allowed to run for a long time. The computation is stopped when the estimated probability $\hat{P}_Q = \Pr(\theta \leq \hat{\theta}_Q)$ reaches within $\pm R$ of the value Q with probability S; that is, $\Pr(Q - R \leq \hat{P}_Q \leq Q + R) = S$. The following *raftery-options* enable you to specify Q, R, S, and a precision level ϵ for the test:

QUANTILE=*value*

Q=*value*
specifies the order (a value between 0 and 1) of the quantile of interest. The default is 0.025.

ACCURACY=*value*

R=*value*
specifies a small positive number as the margin of error for measuring the accuracy of estimation of the quantile. The default is 0.005.

PROBABILITY=*value*

S=*value*
specifies the probability of attaining the accuracy of the estimation of the quantile. The default is 0.95.

EPSILON=*value*

EPS=*value*
specifies the tolerance level (a small positive number) for the stationary test. The default is 0.001.

MARGINLIKE<(NSIM=*number***)>**
evaluates of the logarithm of the marginal likelihood. Two estimates are produced: the cross entropy estimate and the harmonic mean. The cross entropy estimate is based on an importance sampling algorithm. You can specify the number of importance samples in the NSIM=*number* option. By default NSIM=10000. For more information, see the section "Marginal Likelihood" on page 2027.

MAXTUNE=*number*
specifies the maximum number of tuning phases. The default is 24.

MINTUNE=*number*
specifies the minimum number of tuning phases. The default is 2.

NBI=*number*
specifies the number of burn-in iterations before the chains are saved. The default is 1,000.

NMC=*number*
specifies the number of iterations after the burn-in for Metropolis sampling scheme. The default is 1,000.

NMCPRIOR=*number*
specifies the number of samples for the prior predictive analysis when PLOTS(PRIOR)=BAYESPRED is requested. The default is 10,000.

NTRDS=*number*

THREADS=*number*
specifies the number of threads to be used. The number of threads cannot exceed the number of computer cores available. Each core samples the number of iterations that is specified by the NMC option. The default is 1.

NTU=*number*
specifies the number of samples for each tuning phase for Metropolis sampling scheme. The default is 500.

OUTPOST=*SAS-data-set*
names the SAS data set to contain the posterior samples. Alternatively, you can create the output data set by specifying an ODS OUTPUT statement as follows:

ODS OUTPUT POSTERIORSAMPLE = < *SAS-data-set* > **;**

OUTPRIOR=*SAS-data-set*
names the SAS data set to contain the prior samples used to generate the prior predictive analysis when you request the prior predictive plots. Alternatively, you can create the output data set by specifying an ODS OUTPUT statement as follows:

ODS OUTPUT PRIORSAMPLE = < *SAS-data-set* > **;**

PROPCOV=*value*
specifies the method used in constructing the initial covariance matrix for the Metropolis-Hastings algorithm. The QUANEW and NMSIMP methods find numerically approximated covariance matrices at the optimum of the posterior density function with respect to all continuous parameters. The tuning phase starts at the optimized values; in some problems, this can greatly increase convergence performance. If the approximated covariance matrix is not positive definite, then an identity matrix is used instead. You can specify the following *values*:

CONGRA performs a conjugate-gradient optimization.

DBLDOG performs a version of double-dogleg optimization.

NEWRAP performs a Newton-Raphson optimization that combines a line-search algorithm with ridging.

NMSIMP performs a Nelder-Mead simplex optimization.

NRRIDG performs a Newton-Raphson optimization with ridging.

QUANEW performs a quasi-Newton optimization.

TRUREG performs a trust-region optimization.

SAMPLING=*value*

specifies how to sample from the posterior distribution. You can specify the following *values*:

MODELMETROPOLIS

implements a Metropolis sampling scheme on multiple blocks: one block for each model (all the parameters of the model) plus a block for all the correlation parameters across the models.

MULTIMETROPOLIS

implements a Metropolis sampling scheme on a single block that contains all the parameters of the model. SAMPLING=MULTIMETROPOLIS is the default option.

UNIMETROPOLIS

implements a Metropolis sampling scheme on multiple blocks, one for each parameter of the model.

SEED=*number*

specifies an integer seed in the range 1 to $2^{31} - 1$ for the random number generator in the simulation. Specifying a seed enables you to reproduce identical Markov chains for the same specification. If you do not specify the SEED= option, or if you specify a nonpositive seed, a random seed is derived from the time of day.

SIMTIME

prints the time required for the MCMC sampling.

STATISTICS < (*global-options***) > = ALL | NONE |** *keyword* **|** *(keyword-list)*

STATS < (*global-options***) > = ALL | NONE |** *keyword* **|** *(keyword-list)*

controls the number of posterior statistics produced. Specifying STATISTICS=ALL is equivalent to specifying STATISTICS= (CORR COV INTERVAL PRIOR SUMMARY). If you do not want any posterior statistics, specify STATISTICS=NONE. The default is STATISTICS=(SUMMARY INTERVAL). You can specify the following *global-options*:

ALPHA=*numeric-list*

controls the probabilities of the credible intervals. The ALPHA= values must be between 0 and 1. Each ALPHA= value produces a pair of 100(1–ALPHA)% equal-tail and HPD intervals for each parameter. The default is ALPHA=0.05, which yields the 95% credible intervals for each parameter.

PERCENT=*numeric-list*

requests the percentile points of the posterior samples. The PERCENT= values must be between 0 and 100. The default is PERCENT=25, 50, 75, which yields the 25th, 50th, and 75th percentile points, respectively, for each parameter.

You can specify the following *keywords*:

CORR	produces the posterior correlation matrix.
COV	produces the posterior covariance matrix.
INTERVAL	produces equal-tail credible intervals and HPD intervals. The default is to produce the 95% equal-tail credible intervals and 95% HPD intervals, but you can use the ALPHA= *global-option* to request intervals of any probabilities.
NONE	suppresses printing of all summary statistics.
PRIOR	produces a summary table of the prior distributions used in the Bayesian analysis.
SUMMARY	produces the means, standard deviations, and percentile points (25th, 50th, and 75th) for the posterior samples. You can use the global PERCENT= *global-option* to request specific percentile points.

THIN=*number*

THINNING=*number*

controls the thinning of the Markov chain. Only one in every k samples is used when THIN=k, and if NBI=n_0 and NMC=n, the number of samples that are kept is

$$\left[\frac{n_0 + n}{k}\right] - \left[\frac{n_0}{k}\right]$$

where $[a]$ represents the integer part of the number a. The default is THIN=1.

BOUNDS Statement

BOUNDS *bound1* < , *bound2* . . . > ;

The BOUNDS statement imposes simple boundary constraints on the parameter estimates. BOUNDS statement constraints refer to the parameters estimated by the QLIM procedure. Any number of BOUNDS statements can be specified.

Each *bound* is composed of parameters and constants and inequality operators. Parameters associated with regressor variables are referred to by the names of the corresponding regressor variables:

item operator item < *operator item* < *operator item* . . . > >

Each *item* is a constant, the name of a parameter, or a list of parameter names. For more information about how parameters are named in the QLIM procedure, see the section "Naming of Parameters" on page 2036. Each *operator* is '<', '>', '<=', or '>='.

Both the BOUNDS statement and the RESTRICT statement can be used to impose boundary constraints; however, the BOUNDS statement provides a simpler syntax for specifying these kinds of constraints. For more information, see the section "RESTRICT Statement" on page 1983.

The following BOUNDS statement constrains the estimates of the parameters associated with the variable ttime and the variables x1 through x10 to be between 0 and 1. This example illustrates the use of parameter lists to specify boundary constraints.

```
bounds 0 < ttime x1-x10 < 1;
```

The following BOUNDS statement constrains the estimates of the correlation (_RHO) and sigma (_SIGMA) in the bivariate model:

```
bounds _rho >= 0, _sigma.y1 > 1, _sigma.y2 < 5;
```

The BOUNDS statement is not supported if a BAYES statement is also specified. In Bayesian analysis, the restrictions on parameters are usually introduced through the prior distribution.

BY Statement

BY *variables* **;**

A BY statement can be used with PROC QLIM to obtain separate analyses on observations in groups defined by the BY variables.

CLASS Statement

CLASS *variables* **;**

The CLASS statement names the classification variables to be used in the analysis. Classification variables can be either character or numeric.

Class levels are determined from the formatted values of the CLASS variables. Thus, you can use formats to group values into levels. For more information, see the discussion of the FORMAT procedure in *SAS Language Reference: Dictionary*.

ENDOGENOUS Statement

ENDOGENOUS *variables* ∼ *options* **;**

The ENDOGENOUS statement specifies the type of dependent variables that appear on the left-hand side of the equation. Endogenous variables listed refer to the dependent variables that appear on the left-hand side of the equation.

Discrete Variable Options

DISCRETE < (*discrete-options* **) >**
specifies that the endogenous variables in this statement are discrete. Valid *discrete-options* are as follows:

ORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sorting order for the levels of the discrete variables specified in the ENDOGENOUS statement. This ordering determines which parameters in the model correspond to each level in the data. The following table shows how PROC QLIM interprets values of the ORDER= option:

Value of ORDER=	Levels Sorted By
DATA	Order of appearance in the input data set
FORMATTED	Formatted value
FREQ	Descending frequency count; levels with the most observations come first in the order
INTERNAL	Unformatted value

By default, ORDER=FORMATTED. For the values FORMATTED and INTERNAL, the sort order is machine dependent. For more information about sorting order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide*.

DISTRIBUTION=NORMAL | LOGISTIC

DIST=NORMAL | LOGISTIC

D=NORMAL | LOGISTIC

specifies the cumulative distribution function used to model the response probabilities. You can specify the following values:

NORMAL specifies the normal distribution for the probit model.

LOGISTIC specifies the logistic distribution for the logit model.

By default, DISTRIBUTION=NORMAL.

If a multivariate model is specified, logistic distribution is not allowed. Only normal distribution is supported.

Censored Variable Options

CENSORED (*censored-options* **)**

specifies that the endogenous variables in this statement be censored. Valid *censored-options* are as follows:

LB=*value* | *variable*

LOWERBOUND=*value* | *variable*

specifies the lower bound of the censored variables. If *value* is missing or the value in *variable* is missing, no lower bound is set. By default, no lower bound is set.

UB=*value* | *variable*

UPPERBOUND=*value* | *variable*

specifies the upper bound of the censored variables. If *value* is missing or the value in *variable* is missing, no upper bound is set. By default, no upper bound is set.

Truncated Variable Options

TRUNCATED (*truncated-options* **)**

specifies that the endogenous variables in this statement be truncated. Valid *truncated-options* are as follows:

LB=*value* | *variable*
LOWERBOUND=*value* | *variable*
specifies the lower bound of the truncated variables. If *value* is missing or the value in *variable* is missing, no lower bound is set. By default, no lower bound is set.

UB=*value* | *variable*
UPPERBOUND=*value* | *variable*
specifies the upper bound of the truncated variables. If *value* is missing or the value in *variable* is missing, no upper bound is set. By default, no upper bound is set.

Stochastic Frontier Variable Options

FRONTIER <(*frontier-options* **)>**
specifies that the endogenous variable in this statement follow a production or cost frontier. You can specify the following *frontier-options*:

TYPE=HALF | EXPONENTIAL | TRUNCATED
specifies the model type. You can specify the following values:

HALF	specifies a half-normal model.
EXPONENTIAL	specifies an exponential model.
TRUNCATED	specifies a truncated normal model.

PRODUCTION
specifies that the model estimated be a production function.

COST
specifies that the model estimated be a cost function.

If neither the PRODUCTION option nor the COST option is specified, production function is estimated by default.

Selection Options

SELECT (*select-option* **)**
specifies selection criteria for sample selection model. The BAYES statement does not support the SELECT option. The *select-option* specifies the condition for the endogenous variable to be selected. It is written as a variable name, followed by an equality operator (=) or an inequality operator (<, >, <=, >=), followed by a number:

variable operator number

The *variable* is the endogenous variable that the selection is based on. The *operator* can be =, <, >, <= , or >=. Multiple *select-options* can be combined with the logic operators: AND, OR. The following example illustrates the use of the SELECT option:

```
endogenous y1 ~ select(z=0);
endogenous y2 ~ select(z=1 or z=2);
```

The SELECT option can be used together with the DISCRETE, CENSORED, or TRUNCATED option. For example:

```
endogenous y1 ~ select(z=0) discrete;
endogenous y2 ~ select(z=1) censored (lb=0);
endogenous y3 ~ select(z=1 or z=2) truncated (ub=10);
```

For more information about selection models with censoring or truncation, see the section "Selection Models" on page 1997.

Endogeneity and Overidentification Test Options

ENDOTEST (*regressors*)

requests the test of endogeneity for a list of regressors in the model. More specifically, this option tests the null hypothesis that the specified regressors are exogenous. Each of these regressors must also have a model of its own. The former model is considered the structural model, and the latter models are considered reduced form models.

The following example illustrates the use of the ENDOTEST option by testing whether the regressors $y2$ and $y3$ are endogenous in the model for $y1$:

```
proc qlim;
   model y1 = y2 y3 x1;
   model y2 = x1 x2 x3 x4 x5;
   model y3 = x1 x2 x3 x4 x5;
   endogenous y1 ~ endotest(y2 y3);
run;
```

The ENDOTEST option is not available when you specify the SELECT or FRONTIER option. You can specify the ENDOTEST option only once for each ENDOGENOUS statement.

For more information about the test for endogeneity, see the section "Test for Endogeneity" on page 2007.

OVERID (*variables*)

requests the overidentification test for a list of variables. These variables are the overidentifying instrumental variables that you provide from the reduced form models. For more information, see the section "Overidentification Test" on page 2008.

The following example illustrates the use of the OVERID option:

```
proc qlim;
   model y1 = y2 y3 x1;
   model y2 = x1 x2 x3 x4 x5;
   model y3 = x1 x2 x3 x4 x5;
   endogenous y1 ~ overid(y2.x4 y3.x5);
run;
```

The regressors $y2$ and $y3$ in the model for $y1$ are the endogenous variables. Therefore, each of these variables has its own models, which are considered reduced form models. The overidentifying instrumental variables are $x4$ and $x5$. If you specify the OVERID option as

```
endogenous y1 ~ overid(y2.x4 y2.x5);
```

then you consider only the regressor $y2$ to be endogenous, and the model for $y3$ is ignored during the testing process.

The OVERID option is not available when you specify the SELECT or FRONTIER option. You can specify the OVERID option only once for each ENDOGENOUS statement.

FREQ Statement

FREQ *variable* **;**

The FREQ statement identifies a variable that contains the frequency of occurrence of each observation. PROC QLIM treats each observation as if it appears n times, where n is the value of the FREQ variable for the observation. If it is not an integer, the frequency value is truncated to an integer. If the frequency value is less than 1 or missing, the observation is not used in the model fitting. When the FREQ statement is not specified, each observation is assigned a frequency of 1. If you specify more than one FREQ statement, then the first FREQ statement is used.

HETERO Statement

HETERO *dependent variables* ~ *exogenous variables* < / *options* > **;**

The HETERO statement specifies variables that are related to the heteroscedasticity of the residuals and the way these variables are used to model the error variance. The heteroscedastic regression model supported by PROC QLIM is

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$$

$$\epsilon_i \sim \mathrm{N}(0, \sigma_i^2)$$

For more information about the specification of functional forms, see the section "Heteroscedasticity" on page 1994. You can specify the following *options* after a slash (/):

LINK=EXP | LINEAR

specifies the functional form. You can specify the following values:

EXP specifies the exponential link function,

$$\sigma_i^2 = \sigma^2(1 + \exp(\mathbf{z}_i'\boldsymbol{\gamma}))$$

LINEAR specifies the linear link function,

$$\sigma_i^2 = \sigma^2(1 + \mathbf{z}_i'\boldsymbol{\gamma})$$

By default, LINK=EXP.

NOCONST

specifies that there be no constant in the exponential heteroscedasticity model.

$$\sigma_i^2 = \sigma^2 \exp(\mathbf{z}_i' \boldsymbol{\gamma})$$

SQUARE

estimates the model by using the square of linear heteroscedasticity function. For example, you can specify the following heteroscedasticity function:

$$\sigma_i^2 = \sigma^2 (1 + (\mathbf{z}_i' \boldsymbol{\gamma})^2)$$

```
model y = x1 x2 / discrete;
hetero y ~ z1 / link=linear square;
```

The option SQUARE does not apply to exponential heteroscedasticity function because the square of an exponential function of $\mathbf{z}_i' \boldsymbol{\gamma}$ is the same as the exponential of $2\mathbf{z}_i' \boldsymbol{\gamma}$. Hence the only difference is that all $\boldsymbol{\gamma}$ estimates are divided by two.

You can use the HETERO statement within a Bayesian framework, but you should do this carefully because convergence can be slower than in the homoscedastic case. For more information, see the section "Priors for Heteroscedastic Models" on page 2023.

INIT Statement

INIT *initvalue1* < , *initvalue2* . . . > **;**

The INIT statement sets initial values for parameters in the optimization. You can specify any number of INIT statements.

Each *initvalue* is written as a parameter or parameter list, followed by an optional equality operator (=), followed by a number:

parameter **<=>** *number*

If you also specify the BAYES statement, the INIT statement also initializes the Markov chain Monte Carlo (MCMC) algorithm. In particular, the INIT statement does one of the following:

- It initializes the tuning phase (this also includes the PROPCOV option).
- It initializes the sampling phase, if there is no tuning phase.

MODEL Statement

MODEL *dependent-variable* = *regressors* < / *options* > **;**

The MODEL statement specifies the dependent variable and independent regressor variables for the regression model.

You can specify the following *options* after a slash (/):

LIMIT1=ZERO | VARYING
specifies the restriction of the threshold value of the first category when the ordinal probit or logit model is estimated. LIMIT1=ZERO is the default option. When LIMIT1=VARYING is specified, the threshold value is estimated.

NOINT
suppresses the intercept parameter.

Endogenous Variable Options

The endogenous variable options are the same as the options that you can specify in the ENDOGENOUS statement. If you specify an ENDOGENOUS statement, all endogenous options in the MODEL statement are ignored.

Endogeneity and Overidentification Test Options

The endogeneity and overidentification test options are the same as the options that you can specify in the ENDOGENOUS statement. If you specify an ENDOGENOUS statement, all endogeneity and overidentification test options in the MODEL statement are ignored.

BOXCOX Estimation Options

BOXCOX (*option-list* **)**
specifies options that are used for Box-Cox regression or regressor transformation. For example, the Box-Cox regression is specified as

```
model y = x1 x2 / boxcox(y=lambda,x1 x2)
```

PROC QLIM estimates the following Box-Cox regression model:

$$y_i^{(\lambda)} = \beta_0 + \beta_1 x_{1i}^{(\lambda_2)} + \beta_2 x_{2i}^{(\lambda_2)} + \epsilon_i$$

The *option-list* takes the form *variable-list* < = *varname* > separated by commas. The *variable-list* specifies that the list of variables have the same Box-Cox transformation; *varname* specifies the name of this Box-Cox coefficient. If *varname* is not specified, the coefficient is called _Lambda*i*, where *i* increments sequentially.

Variable Selection Options

SELECTVAR <=(*selectvar-option***)>**
enables variable selection. The *selectvar-option* specifies a variable selection method based on an information criterion. For more information, see the section "Variable Selection" on page 2001. You can specify the following *selectvar-options*:

DIRECTION=FORWARD | BACKWARD
specifies the searching algorithm to use in the variable selection method. By default, DIRECTION=FORWARD.

CRITER=AIC | SBC
specifies the information criterion to use for the variable selection. By default, CRITER=AIC.

MAXSTEPS=*value*
specifies the maximum number of steps that are allowed in the search algorithm. The default is 100.

LSTOP=*value*
specifies the stopping criterion. The *value* represents the percentage of decrease or increase in the AIC or SBC that is required for the algorithm to proceed; it must be a positive number less than 1. The default is 0.

RETAIN(*regressors***)**
specifies a list of regressors that are to be retained in any model that the variable selection process considers.

The following rules apply to how regressors are handled when you specify more than one MODEL statement and you use the SELECTVAR option:

- If you do not specify the SELECTVAR option in a particular MODEL statement, then all regressors in the original model are included in any model that the variable selection algorithm considers. In other words, omitting the SELECTVAR option is equivalent to providing the option: SELECTVAR=(RETAIN(*all-regressors*)).
- If you specify the SELECTVAR option without any =(*option*) clause in a MODEL statement, then all regressors in that model (other than the intercept, if present) are eligible for potential exclusion as the variable selection process is executed.

The following example specifies 10 possible regressor candidates, 5 of which are selected using the AIC:

```
proc qlim data=one;
   model y = x1-x10 /selectvar=(direction=forward criter=AIC maxsteps=5);
run;
```

NLOPTIONS Statement

NLOPTIONS < *options* > **;**

PROC QLIM uses the nonlinear optimization (NLO) subsystem to perform nonlinear optimization tasks. For a list of all the options of the NLOPTIONS statement, see Chapter 6, "Nonlinear Optimization Methods."

OUTPUT Statement

OUTPUT < **OUT**=*SAS-data-set* > < *output-options* > **;**

The OUTPUT statement creates a new SAS data set containing all variables in the input data set and, optionally, the estimates of $\mathbf{x}'\boldsymbol{\beta}$, predicted value, residual, marginal effects, probability, standard deviation

of the error, expected value, conditional expected value, technical efficiency measures, and inverse Mills ratio. When the response values are missing for the observation, all output estimates except residual are still computed as long as none of the explanatory variables is missing. This enables you to compute these statistics for prediction. You can specify only one OUTPUT statement.

Details about the specifications in the OUTPUT statement are as follows:

CONDITIONAL
outputs estimates of conditional expected values of continuous endogenous variables.

ERRSTD
outputs estimates of σ_j, the standard deviation of the error term.

EXPECTED
outputs estimates of expected values of continuous endogenous variables.

MARGINAL
outputs marginal effects.

MILLS
outputs estimates of inverse Mills ratios of censored or truncated continuous, binary discrete, and selection endogenous variables.

OUT=*SAS-data-set*
names the output data set.

PREDICTED
outputs estimates of predicted endogenous variables.

PROB
outputs estimates of probability of discrete endogenous variables taking the current observed responses.

PROBALL
outputs estimates of probability of discrete endogenous variables for all possible responses.

RESIDUAL
outputs estimates of residuals of continuous endogenous variables.

XBETA
outputs estimates of $\mathbf{x}'\boldsymbol{\beta}$.

TE1
outputs estimates of technical efficiency for each producer in the stochastic frontier model suggested by Battese and Coelli (1988).

TE2
outputs estimates of technical efficiency for each producer in the stochastic frontier model suggested by Jondrow et al. (1982).

PRIOR Statement

PRIOR *parameter-list* ∼ *distribution* **;**

PRIOR _REGRESSORS ;

The PRIOR statement specifies the prior distribution of the model parameters. You must specify a single parameter or a list of parameters, a tilde ∼, and then a distribution with its parameters. Alternately, you can specify the special keyword REGRESSORS to select all the parameters used in the linear regression component of the model. Multiple PRIOR statements are allowed.

You can specify the following *distributions*:

NORMAL(MEAN=μ, VAR=σ^2)
specifies a normal distribution with parameters MEAN and VAR.

GAMMA(SHAPE=*a*, SCALE=*b*)
specifies a gamma distribution with parameters SHAPE and SCALE.

SQGAMMA(SHAPE=*a*, SCALE=*b*)
specifies a square root gamma distribution with parameters SHAPE and SCALE.

IGAMMA(SHAPE=*a*, SCALE=*b*)
specifies an inverse gamma distribution with parameters SHAPE and SCALE.

SQIGAMMA(SHAPE=*a*, SCALE=*b*)
specifies a square root inverse gamma distribution with parameters SHAPE and SCALE.

UNIFORM(MIN=*m*, MAX=*M*)
specifies a uniform distribution that is defined between MIN and MAX.

BETA(SHAPE1=*a*, SHAPE2=*b*, MIN=*m*, MAX=*M*)
specifies a beta distribution with parameters SHAPE1 and SHAPE2 and defined between MIN and MAX.

T(LOCATION=μ, DF=ν)
specifies a noncentral t distribution with DF degrees of freedom and location parameter equal to LOCATION.

For information about how to specify *distributions*, see the section "Standard Distributions" on page 2029.

RANDOM Statement (Experimental)

RANDOM *regressors* < / *options* > **;**

The RANDOM statement defines the regressors of the model, including the intercept, that have random coefficients in a random-parameters model. If you have a panel data set, you can use the RANDOM statement to estimate random-parameters models that include binomial probit, binomial logit, ordinal probit, ordinal logit, linear regression, Tobit, truncated regression, and stochastic frontier models. You do not have to have the observations collected in a panel data setting to model the parameter heterogeneity. Random-parameters models can also be applied to cross-sectional data.

If you only have a group heterogeneity in your error term, or individual specific constant terms as randomly distributed across the groups, then you have a random-effects model and in this case you specify *regressors* as INTERCEPT (or INT) only.

You can specify only a single RANDOM statement, and if you specify a RANDOM statement, you can specify only one MODEL statement. The RANDOM statement is not supported if a BAYES statement is also specified.

You can specify the following *options* after a slash (/).

SUBJECT=*variable*

S=*variable*

determines the variable that specifies the ID of the individuals or groups across which the parameter heterogeneity occurs. In panel data, the *variable* identifies the cross-sectional units. For example, in panel data, the *variable* might be household or country.

If you do not specify this option, then *variable* is assumed to have a single realization; that is, there is no variation in the random effects. You should specify this option in order to have a true random-parameters model.

The following statement illustrates this option:

```
random int / subject=id;
```

METHOD=*method-options*

M=*method-options*

specifies the method of approximation to the integral that appears in the likelihood function. For more information about the integral and the integration methods, see the section "Random-Parameters Models and Panel Data Analysis" on page 2008 and its subsections.

You can specify the following *method-options*:

HALTON < (*halton-options***) >**

HALT < (*halton-options***) >**

QMC < (*halton-options***) >**

requests a quasi–Monte Carlo integration method that uses the Halton sequences that are defined by the prime numbers starting from 2. For information about how this series is generated, see the section "QMC Method Using the Halton Sequence" on page 2014.

You can specify the following *halton-options*:

NDRAW=*value*

determines the number of elements that the Halton series has for each unique value of the subject variable. Therefore, the total number of elements in the Halton sequence is *value* times the number of unique values of the *variable* that you specify in the SUBJECT= option. For more information, see the section "QMC Method Using the Halton Sequence" on page 2014.

The default value of the NDRAW= option is the number of unique values of the *variable* that you specify in the SUBJECT= option. For example, if you have a panel data set, the total number of terms in the Halton sequence is the square of the number of cross sections.

START=*value*
specifies the starting point of the Halton sequence, where *value* must be a positive integer. When you specify this option, *value*–1 extra draws are created and the initial *value*–1 elements are discarded. By default, START=11.

The following statement estimates a random-effects model and requests a Halton sequence that has 100 draws for each country and does not discard any draws:

```
random int / subject=country method=halton(ndraw=100 start=1);
```

The following statements estimate a random-parameters probit model by specifying a random intercept and unobserved heterogeneity in the coefficients for x1 and x2. The statements request 500 Halton draws and discard the first 50 elements for each of the three sequences.

```
proc qlim data=a;
   model y = x1 x2 x3 / discrete;
   random int x1 x2 / subject=id
                      method=halton(ndraw=500 start=51);
run;
```

HERMITE < (QPOINTS=*value*) >
HERM < (QPOINTS=*value*) >
GAUSS < (QPOINTS=*value*) >
requests the Gauss-Hermite quadrature integration method. You can use this method if your model has only one random parameter—that is, if you have a random-effects model or if your model has a single random coefficient. For more information about this method, see the section "Approximation by Hermite Quadrature" on page 2014.

QPOINTS=*value* specifies the number of quadrature points to be used during evaluation of the integral. By default, QPOINTS=20.

The following statements illustrate this option for a random-effects model and a random-parameters model with a single random coefficient on x1:

```
random int / subject=states method=hermite(qpoints=4);

random x1 / subject=id method=hermite(qpoints=32);
```

SIMULATION < (*simulation-options*) >
SIM < (*simulation-options*) >
requests Monte Carlo simulation as the method of integration. For more information, see the section "Monte Carlo Integration" on page 2013.

You can specify the following *simulation-options*:

NDRAW=*value*
specifies the number of draws for the simulation. You can also specify the number of draws in the NDRAW= option in the PROC QLIM statement. If you specify this option in both statements, PROC QLIM uses the *value* in the RANDOM statement. If you do not specify this option in either statement, the default value is set to $N^{3/2}$, where N is the number of

unique values of the subject variable. For example, for a panel data set, N is the number of cross sections.

SEED=*value*
specifies the seed of the random draws, where *value* must be less than $2^{31} - 1$. You can also specify the seed in the SEED= option in the PROC QLIM statement. If you specify this option in both statements, PROC QLIM uses the *value* in the RANDOM statement. If you do not specify a seed, or if you specify a *value* less than or equal to zero, the seed is generated randomly.

The following statement illustrates this option:

```
random int x1 / subject=id method=simulation(ndraw=1000 seed=12345);
```

By default, METHOD=HALTON.

NOCORRELATION
NOCORR
requests that the random parameters be uncorrelated with one another. If you specify this option, only the diagonal elements of the covariance matrix of the random parameters are estimated.

RESTRICT Statement

RESTRICT *restriction1* < , *restriction2 . . .* > **;**

The RESTRICT statement is used to impose linear restrictions on the parameter estimates. Any number of RESTRICT statements can be specified, but the number of restrictions imposed is limited by the number of regressors.

Each *restriction* is written as an expression, followed by an equality operator (=) or an inequality operator (<, >, <=, >=), followed by a second expression:

expression operator expression

The *operator* can be =, <, >, <= , or >=. The operator and second expression are optional.

Restriction expressions can be composed of parameter names, multiplication ($*$), addition ($+$) and substitution ($-$) operators, and constants. Parameters named in restriction expressions must be among the parameters estimated by the model. Parameters associated with a regressor variable are referred to by the name of the corresponding regressor variable. The restriction expressions must be a linear function of the parameters.

The following is an example of the use of the RESTRICT statement:

```
proc qlim data=one;
   model y = x1-x10 / discrete;
   restrict x1*2 <= x2 + x3;
run;
```

The RESTRICT statement can also be used to impose cross-equation restrictions in multivariate models. The following RESTRICT statement imposes an equality restriction on coefficients of $x1$ in equation $y1$ and $x1$ in equation $y2$:

```
proc qlim data=one;
   model y1 = x1-x10;
   model y2 = x1-x4;
   endogenous y1 y2 ~ discrete;
   restrict y1.x1=y2.x1;
run;
```

The RESTRICT statement is not supported if a BAYES statement is also specified. In Bayesian analysis, the restrictions on parameters are usually introduced through the prior distribution.

TEST Statement

<'label':> **TEST** *<'string':> equation [,equation. . .] / options* **;**

The TEST statement performs Wald, Lagrange multiplier, and likelihood ratio tests of linear hypotheses about the regression parameters in the preceding MODEL statement. Each equation specifies a linear hypothesis to be tested. All hypotheses in one TEST statement are tested jointly. Variable names in the equations must correspond to regressors in the preceding MODEL statement, and each name represents the coefficient of the corresponding regressor. The keyword INTERCEPT refers to the coefficient of the intercept.

You cannot specify both the TEST statement and the BAYES statement.

You can specify the following *options* after a slash (/):

ALL
requests Wald, Lagrange multiplier, and likelihood ratio tests.

WALD
requests the Wald test.

LM
requests the Lagrange multiplier test.

LR
requests the likelihood ratio test.

The following illustrates the use of the TEST statement:

```
proc qlim;
   model y = x1 x2 x3;
   test x1 = 0, x2 * .5 + 2 * x3 = 0;
   test_int: test intercept = 0, x3 = 0;
run;
```

The first test investigates the joint hypothesis that

$$\beta_1 = 0$$

and

$$0.5\beta_2 + 2\beta_3 = 0$$

In case there is more than one MODEL statement in one QLIM procedure, then TEST statement is capable of testing cross-equation restrictions. Each parameter reference should be preceded by the name of the dependent variable of the particular model and the dot sign. For example:

```
proc qlim;
   model y1 = x1 x2 x3;
   model y2 = x3 x5 x6;
   test y1.x1 + y2.x6 = 1;
run;
```

This cross-equation test investigates the null hypothesis that

$$\beta_{1,1} + \beta_{2,3} = 1$$

in the system of equations

$$\begin{aligned} y_{1,i} &= \alpha_1 + \beta_{1,1}x_{1,i} + \beta_{1,2}x_{2,i} + \beta_{1,3}x_{3,i} \\ y_{2,i} &= \alpha_2 + \beta_{2,1}x_{3,i} + \beta_{2,2}x_{5,i} + \beta_{2,3}x_{6,i} \end{aligned}$$

Only linear equality restrictions and tests are permitted in PROC QLIM. Tests expressions can be composed only of algebraic operations involving the addition symbol (+), subtraction symbol (-), and multiplication symbol (*).

The TEST statement accepts labels that are reproduced in the printed output. TEST statement can be labeled in two ways. A TEST statement can be preceded by a label followed by a colon. Alternatively, the keyword TEST can be followed by a quoted string. If both are present, PROC QLIM uses the label preceding the colon. In the event no label is present, PROC QLIM automatically labels the tests.

WEIGHT Statement

WEIGHT *variable* < / *option* > **;**

The WEIGHT statement specifies a variable to supply weighting values to use for each observation in estimating parameters. The log likelihood for each observation is multiplied by the corresponding weight variable value.

If the weight of an observation is nonpositive, that observation is not used in the estimation.

You can specify the following *option* after a slash (/):

NONORMALIZE

specifies that the weights are required to be used as is. When this option is not specified, the weights are normalized so that they add up to the actual sample size. Weights w_i are normalized by multiplying them by $\frac{n}{\sum_{i=1}^{n} w_i}$, where n is the sample size.

Details: QLIM Procedure

Ordinal Discrete Choice Modeling

Binary Probit and Logit Model

The binary choice model is

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$$

where value of the latent dependent variable, y_i^*, is observed only as follows:

$$\begin{aligned} y_i &= 1 \quad \text{if } y_i^* > 0 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The disturbance, ϵ_i, of the probit model has standard normal distribution with the distribution function (CDF)

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt$$

The disturbance of the logit model has standard logistic distribution with the CDF

$$\Lambda(x) = \frac{\exp(x)}{1+\exp(x)} = \frac{1}{1+\exp(-x)}$$

The binary discrete choice model has the following probability that the event $\{y_i = 1\}$ occurs:

$$P(y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}) = \begin{cases} \Phi(\mathbf{x}_i'\boldsymbol{\beta}) & \text{(probit)} \\ \Lambda(\mathbf{x}_i'\boldsymbol{\beta}) & \text{(logit)} \end{cases}$$

The log-likelihood function is

$$\ell = \sum_{i=1}^{N} \left\{ y_i \log[F(\mathbf{x}_i'\boldsymbol{\beta})] + (1-y_i)\log[1-F(\mathbf{x}_i'\boldsymbol{\beta})] \right\}$$

where the CDF $F(x)$ is defined as $\Phi(x)$ for the probit model while $F(x) = \Lambda(x)$ for logit. The first-order derivatives of the logit model are

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} (y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i$$

The probit model has more complicated derivatives

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left\{ \frac{(2y_i-1)\phi\left[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}\right]}{\Phi\left[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}\right]} \right\} \mathbf{x}_i = \sum_{i=1}^{N} r_i \mathbf{x}_i$$

where

$$r_i = \frac{(2y_i-1)\phi\left[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}\right]}{\Phi\left[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}\right]}$$

Note that the logit maximum likelihood estimates are $\frac{\pi}{\sqrt{3}}$ times greater than probit maximum likelihood estimates, since the probit parameter estimates, $\boldsymbol{\beta}$, are standardized, and the error term with logistic distribution has a variance of $\frac{\pi^2}{3}$.

Ordinal Probit/Logit

When the dependent variable is observed in sequence with M categories, binary discrete choice modeling is not appropriate for data analysis. McKelvey and Zavoina (1975) proposed the ordinal (or ordered) probit model.

Consider the regression equation

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$$

where error disturbances, ϵ_i, have the distribution function F. The unobserved continuous random variable, y_i^*, is identified as M categories. Suppose there are $M+1$ real numbers, $\mu_0, \ldots, \mu_M$, where $\mu_0 = -\infty$, $\mu_1 = 0$, $\mu_M = \infty$, and $\mu_0 \leq \mu_1 \leq \cdots \leq \mu_M$. Define

$$R_{i,j} = \mu_j - \mathbf{x}_i'\boldsymbol{\beta}$$

The probability that the unobserved dependent variable is contained in the jth category can be written as

$$P[\mu_{j-1} < y_i^* \leq \mu_j] = F(R_{i,j}) - F(R_{i,j-1})$$

The log-likelihood function is

$$\ell = \sum_{i=1}^{N}\sum_{j=1}^{M} d_{ij} \log\left[F(R_{i,j}) - F(R_{i,j-1})\right]$$

where

$$d_{ij} = \begin{cases} 1 & \text{if} \quad \mu_{j-1} < y_i \leq \mu_j \\ 0 & \text{otherwise} \end{cases}$$

The first derivatives are written as

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{j=1}^{M} d_{ij}\left[\frac{f(R_{i,j-1}) - f(R_{i,j})}{F(R_{i,j}) - F(R_{i,j-1})}\mathbf{x}_i\right]$$

$$\frac{\partial \ell}{\partial \mu_k} = \sum_{i=1}^{N}\sum_{j=1}^{M} d_{ij}\left[\frac{\delta_{j,k} f(R_{i,j}) - \delta_{j-1,k} f(R_{i,j-1})}{F(R_{i,j}) - F(R_{i,j-1})}\right]$$

where $f(x) = \frac{dF(x)}{dx}$ and $\delta_{j,k} = 1$ if $j = k$, and $\delta_{j,k} = 0$ otherwise. When the ordinal probit is estimated, it is assumed that $F(R_{i,j}) = \Phi(R_{i,j})$. The ordinal logit model is estimated if $F(R_{i,j}) = \Lambda(R_{i,j})$. The first threshold parameter, μ_1, is estimated when the LIMIT1=VARYING option is specified. By default (LIMIT1=ZERO), so that $M-2$ threshold parameters $(\mu_2, \ldots, \mu_{M-1})$ are estimated.

The ordered probit models are analyzed by Aitchison and Silvey (1957), and Cox (1970) discussed ordered response data by using the logit model. They defined the probability that y_i^* belongs to jth category as

$$P[\mu_{j-1} < y_i \leq \mu_j] = F(\mu_j + \mathbf{x}_i'\boldsymbol{\theta}) - F(\mu_{j-1} + \mathbf{x}_i'\boldsymbol{\theta})$$

where $\mu_0 = -\infty$ and $\mu_M = \infty$. Therefore, the ordered response model analyzed by Aitchison and Silvey can be estimated if the LIMIT1=VARYING option is specified. Note that $\boldsymbol{\theta} = -\boldsymbol{\beta}$.

Goodness-of-Fit Measures

The goodness-of-fit measures discussed in this section apply only to discrete dependent variable models.

McFadden (1974) suggested a likelihood ratio index that is analogous to the R^2 in the linear regression model,

$$R^2_M = 1 - \frac{\ln L}{\ln L_0}$$

where L is the value of the maximum likelihood function and L_0 is the value of a likelihood function when regression coefficients except an intercept term are zero. It can be shown that L_0 can be written as

$$L_0 = \sum_{j=1}^{M} N_j \ln(\frac{N_j}{N})$$

where N_j is the number of responses in category j.

Estrella (1998) proposes the following requirements for a goodness-of-fit measure to be desirable in discrete choice modeling:

- The measure must take values in [0, 1], where 0 represents no fit and 1 corresponds to perfect fit.
- The measure should be directly related to the valid test statistic for significance of all slope coefficients.
- The derivative of the measure with respect to the test statistic should comply with corresponding derivatives in a linear regression.

Estrella's (1998) measure is written

$$R^2_{E1} = 1 - \left(\frac{\ln L}{\ln L_0}\right)^{-\frac{2}{N}\ln L_0}$$

An alternative measure suggested by Estrella (1998) is

$$R^2_{E2} = 1 - [(\ln L - K)/\ln L_0]^{-\frac{2}{N}\ln L_0}$$

where $\ln L_0$ is computed with null slope parameter values, N is the number observations used, and K represents the number of estimated parameters.

Other goodness-of-fit measures are summarized as follows,

$$R^2_{CU1} = 1 - \left(\frac{L_0}{L}\right)^{\frac{2}{N}} \quad \text{(Cragg-Uhler 1)}$$

$$R^2_{CU2} = \frac{1 - (L_0/L)^{\frac{2}{N}}}{1 - L_0^{\frac{2}{N}}} \quad \text{(Cragg-Uhler 2)}$$

$$R^2_A = \frac{2(\ln L - \ln L_0)}{2(\ln L - \ln L_0) + N} \quad \text{(Aldrich-Nelson)}$$

$$R^2_{VZ} = R^2_A \frac{2\ln L_0 - N}{2\ln L_0} \quad \text{(Veall-Zimmermann)}$$

$$R^2_{MZ} = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}}_i)^2}{N + \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}}_i)^2} \quad \text{(McKelvey-Zavoina)}$$

where $\hat{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ and $\bar{\hat{y}}_i = \sum_{i=1}^{N} \hat{y}_i / N$.

Limited Dependent Variable Models

Censored Regression Models

When the dependent variable is censored, values in a certain range are all transformed to a single value. For example, the standard tobit model can be defined as

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$$

$$y_i = \begin{cases} y_i^* & \text{if} y_i^* > 0 \\ 0 & \text{if} y_i^* \le 0 \end{cases}$$

where $\epsilon_i \sim \text{iid} N(0, \sigma^2)$. The log-likelihood function of the standard censored regression model is

$$\ell = \sum_{i \in \{y_i = 0\}} \ln[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)] + \sum_{i \in \{y_i > 0\}} \ln\left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma\right]$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution and $\phi(\cdot)$ is the probability density function of the standard normal distribution.

The tobit model can be generalized to handle observation-by-observation censoring. The censored model on both of the lower and upper limits can be defined as

$$y_i = \begin{cases} R_i & \text{if } y_i^* \ge R_i \\ y_i^* & \text{if } L_i < y_i^* < R_i \\ L_i & \text{if } y_i^* \le L_i \end{cases}$$

The log-likelihood function can be written as

$$\begin{aligned} \ell \quad = \quad & \sum_{i \in \{L_i < y_i < R_i\}} \ln\left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma\right] + \sum_{i \in \{y_i = R_i\}} \ln\left[\Phi(-\frac{R_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})\right] + \\ & \sum_{i \in \{y_i = L_i\}} \ln\left[\Phi(\frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})\right] \end{aligned}$$

Log-likelihood functions of the lower- or upper-limit censored model are easily derived from the two-limit censored model. The log-likelihood function of the lower-limit censored model is

$$\ell = \sum_{i \in \{y_i > L_i\}} \ln\left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma\right] + \sum_{i \in \{y_i = L_i\}} \ln\left[\Phi(\frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})\right]$$

The log-likelihood function of the upper-limit censored model is

$$\ell = \sum_{i \in \{y_i < R_i\}} \ln\left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma\right] + \sum_{i \in \{y_i = R_i\}} \ln\left[1 - \Phi(\frac{R_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})\right]$$

Types of Tobit Models

Amemiya (1984) classified Tobit models into five types based on characteristics of the likelihood function. For notational convenience, let P denote a distribution or density function, y^*_{ji} is assumed to be normally distributed with mean $\mathbf{x}'_{ji}\boldsymbol{\beta}_j$ and variance σ^2_j.

Type 1 Tobit

The Type 1 Tobit model was already discussed in the preceding section.

$$\begin{aligned} y^*_{1i} &= \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i} \\ y_{1i} &= y^*_{1i} \quad \text{if } y^*_{1i} > 0 \\ &= 0 \quad \text{if } y^*_{1i} \le 0 \end{aligned}$$

The likelihood function is characterized as $P(y_1 < 0)P(y_1)$.

Type 2 Tobit

The Type 2 Tobit model is defined as

$$\begin{aligned} y^*_{1i} &= \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i} \\ y^*_{2i} &= \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + u_{2i} \\ y_{1i} &= 1 \quad \text{if } y^*_{1i} > 0 \\ &= 0 \quad \text{if } y^*_{1i} \le 0 \\ y_{2i} &= y^*_{2i} \quad \text{if } y^*_{1i} > 0 \\ &= 0 \quad \text{if } y^*_{1i} \le 0 \end{aligned}$$

where $(u_{1i}, u_{2i}) \sim N(0, \Sigma)$. The likelihood function is described as $P(y_1 < 0)P(y_1 > 0, y_2)$.

Type 3 Tobit

The Type 3 Tobit model is different from the Type 2 Tobit in that y^*_{1i} of the Type 3 Tobit is observed when $y^*_{1i} > 0$.

$$\begin{aligned} y^*_{1i} &= \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + u_{1i} \\ y^*_{2i} &= \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + u_{2i} \\ y_{1i} &= y^*_{1i} \quad \text{if } y^*_{1i} > 0 \\ &= 0 \quad \text{if } y^*_{1i} \le 0 \\ y_{2i} &= y^*_{2i} \quad \text{if } y^*_{1i} > 0 \\ &= 0 \quad \text{if } y^*_{1i} \le 0 \end{aligned}$$

where $(u_{1i}, u_{2i})' \sim \text{iid}N(0, \Sigma)$.

The likelihood function is characterized as $P(y_1 < 0)P(y_1, y_2)$.

Type 4 Tobit

The Type 4 Tobit model consists of three equations,

$$\begin{aligned}
y_{1i}^{*} &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + u_{1i} \\
y_{2i}^{*} &= \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + u_{2i} \\
y_{3i}^{*} &= \mathbf{x}_{3i}'\boldsymbol{\beta}_3 + u_{3i} \\
y_{1i} &= y_{1i}^{*} \quad \text{if } y_{1i}^{*} > 0 \\
&= 0 \quad \text{if } y_{1i}^{*} \leq 0 \\
y_{2i} &= y_{2i}^{*} \quad \text{if } y_{1i}^{*} > 0 \\
&= 0 \quad \text{if } y_{1i}^{*} \leq 0 \\
y_{3i} &= y_{3i}^{*} \quad \text{if } y_{1i}^{*} \leq 0 \\
&= 0 \quad \text{if } y_{1i}^{*} > 0
\end{aligned}$$

where $(u_{1i}, u_{2i}, u_{3i})' \sim \text{iid}N(0, \Sigma)$. The likelihood function of the Type 4 Tobit model is characterized as $P(y_1 < 0, y_3)P(y_1, y_2)$.

Type 5 Tobit

The Type 5 Tobit model is defined as follows,

$$\begin{aligned}
y_{1i}^{*} &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + u_{1i} \\
y_{2i}^{*} &= \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + u_{2i} \\
y_{3i}^{*} &= \mathbf{x}_{3i}'\boldsymbol{\beta}_3 + u_{3i} \\
y_{1i} &= 1 \quad \text{if } y_{1i}^{*} > 0 \\
&= 0 \quad \text{if } y_{1i}^{*} \leq 0 \\
y_{2i} &= y_{2i}^{*} \quad \text{if } y_{1i}^{*} > 0 \\
&= 0 \quad \text{if } y_{1i}^{*} \leq 0 \\
y_{3i} &= y_{3i}^{*} \quad \text{if } y_{1i}^{*} \leq 0 \\
&= 0 \quad \text{if } y_{1i}^{*} > 0
\end{aligned}$$

where $(u_{1i}, u_{2i}, u_{3i})'$ are from iid trivariate normal distribution. The likelihood function of the Type 5 Tobit model is characterized as $P(y_1 < 0, y_3)P(y_1 > 0, y_2)$.

Code examples for these models can be found in "Example 28.6: Types of Tobit Models" on page 2050.

Truncated Regression Models

In a truncated model, the observed sample is a subset of the population where the dependent variable falls in a certain range. For example, when neither a dependent variable nor exogenous variables are observed for $y_i^{*} < 0$, the truncated regression model can be specified.

$$\ell = \sum_{i \in \{y_i \geq 0\}} \left\{ -\ln \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma) + \ln \left[\frac{\phi((y_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma)}{\sigma} \right] \right\}$$

Two-limit truncation model is defined as

$$y_i = y_i^{*} \text{if } L_i \leq y_i^{*} \leq R_i$$

The log-likelihood function of the two-limit truncated regression model is

$$\ell = \sum_{i=1}^{N} \left\{ \ln \left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma \right] - \ln \left[\Phi(\frac{R_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}) - \Phi(\frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}) \right] \right\}$$

The log-likelihood functions of the lower- and upper-limit truncation model are

$$\ell = \sum_{i=1}^{N} \left\{ \ln \left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma \right] - \ln \left[1 - \Phi(\frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}) \right] \right\} \quad \text{(lower)}$$

$$\ell = \sum_{i=1}^{N} \left\{ \ln \left[\phi(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma})/\sigma \right] - \ln \left[\Phi(\frac{R_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}) \right] \right\} \quad \text{(upper)}$$

Stochastic Frontier Production and Cost Models

Stochastic frontier production models were first developed by Aigner, Lovell, and Schmidt (1977); Meeusen and van den Broeck (1977). Specification of these models allows for random shocks of the production or cost but also includes a term for technological or cost inefficiency. Assuming that the production function takes a log-linear Cobb-Douglas form, the stochastic frontier production model can be written as

$$ln(y_i) = \beta_0 + \sum_n \boldsymbol{\beta}_n \ln(x_{ni}) + \epsilon_i$$

where $\epsilon_i = v_i - u_i$. The v_i term represents the stochastic error component and u_i is the nonnegative, technology inefficiency error component. The v_i error component is assumed to be distributed iid normal and independently from u_i. Given that $u_i > 0$, the error term, ϵ_i, is negatively skewed and represents technology inefficiency. For the stochastic frontier cost model, $\epsilon_i = v_i + u_i$. The v_i term represents the stochastic error component and u_i is the nonnegative, cost inefficiency error component. Given that $u_i > 0$, the error term, ϵ_i, is positively skewed and represents cost inefficiency. PROC QLIM models the u_i error component as a half normal, exponential, or truncated normal distribution.

The Normal–Half Normal Model

In case of the normal–half normal model, v_i is iid $N(0, \sigma_v^2)$, u_i is iid $N^+(0, \sigma_u^2)$ with v_i and u_i independent of each other. Given the independence of error terms, the joint density of v and u can be written as

$$f(u, v) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left\{ -\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2} \right\}$$

Substituting $v = \epsilon + u$ into the preceding equation gives

$$f(u, \epsilon) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left\{ -\frac{u^2}{2\sigma_u^2} - \frac{(\epsilon + u)^2}{2\sigma_v^2} \right\}$$

Integrating u out to obtain the marginal density function of ϵ results in the form

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(u,\epsilon)du \\ &= \frac{2}{\sqrt{2\pi}\sigma}\left[1-\Phi\left(\frac{\epsilon\lambda}{\sigma}\right)\right]\exp\left\{-\frac{\epsilon^2}{2\sigma^2}\right\} \\ &= \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(-\frac{\epsilon\lambda}{\sigma}\right) \end{aligned}$$

where $\lambda = \sigma_u/\sigma_v$ and $\sigma = \sqrt{\sigma_u^2+\sigma_v^2}$.

In the case of a stochastic frontier cost model, $v = \epsilon - u$ and

$$f(\epsilon) = \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(\frac{\epsilon\lambda}{\sigma}\right)$$

The log-likelihood function for the production model with N producers is written as

$$\ln L = constant - N\ln\sigma + \sum_i \ln\Phi\left(-\frac{\epsilon_i\lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_i \epsilon_i^2$$

The Normal-Exponential Model

Under the normal-exponential model, v_i is iid $N(0,\sigma_v^2)$ and u_i is iid exponential with scale parameter σ_u. Given the independence of error term components u_i and v_i, the joint density of v and u can be written as

$$f(u,v) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v}\exp\left\{-\frac{u}{\sigma_u}-\frac{v^2}{2\sigma_v^2}\right\}$$

The marginal density function of ϵ for the production function is

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(u,\epsilon)du \\ &= \left(\frac{1}{\sigma_u}\right)\Phi\left(-\frac{\epsilon}{\sigma_v}-\frac{\sigma_v}{\sigma_u}\right)\exp\left\{\frac{\epsilon}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}\right\} \end{aligned}$$

and the marginal density function for the cost function is equal to

$$f(\epsilon) = \left(\frac{1}{\sigma_u}\right)\Phi\left(\frac{\epsilon}{\sigma_v}-\frac{\sigma_v}{\sigma_u}\right)\exp\left\{-\frac{\epsilon}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}\right\}$$

The log-likelihood function for the normal-exponential production model with N producers is

$$\ln L = constant - N\ln\sigma_u + N\left(\frac{\sigma_v^2}{2\sigma_u^2}\right) + \sum_i\frac{\epsilon_i}{\sigma_u} + \sum_i \ln\Phi\left(-\frac{\epsilon_i}{\sigma_v}-\frac{\sigma_v}{\sigma_u}\right)$$

The Normal–Truncated Normal Model

The normal–truncated normal model is a generalization of the normal–half normal model by allowing the mean of u_i to differ from zero. Under the normal–truncated normal model, the error term component v_i is iid $N(0, \sigma_v^2)$ and u_i is iid $N^+(\mu, \sigma_u^2)$. The joint density of v_i and u_i can be written as

$$f(u,v) = \frac{1}{2\pi\sigma_u\sigma_v\Phi\left(\mu/\sigma_u\right)} \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}$$

The marginal density function of ϵ for the production function is

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(u,\epsilon)du \\ &= \frac{1}{\sqrt{2\pi}\sigma\Phi\left(\mu/\sigma_u\right)}\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\epsilon\lambda}{\sigma}\right)\exp\left\{-\frac{(\epsilon+\mu)^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sigma}\phi\left(\frac{\epsilon+\mu}{\sigma}\right)\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\epsilon\lambda}{\sigma}\right)\left[\Phi\left(\frac{\mu}{\sigma_u}\right)\right]^{-1} \end{aligned}$$

and the marginal density function for the cost function is

$$f(\epsilon) = \frac{1}{\sigma}\phi\left(\frac{\epsilon-\mu}{\sigma}\right)\Phi\left(\frac{\mu}{\sigma\lambda} + \frac{\epsilon\lambda}{\sigma}\right)\left[\Phi\left(\frac{\mu}{\sigma_u}\right)\right]^{-1}$$

The log-likelihood function for the normal–truncated normal production model with N producers is

$$\begin{aligned} \ln L &= constant - N\ln\sigma - N\ln\Phi\left(\frac{\mu}{\sigma_u}\right) + \sum_i \ln\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\epsilon_i\lambda}{\sigma}\right) \\ &\quad -\frac{1}{2}\sum_i\left(\frac{\epsilon_i+\mu}{\sigma}\right)^2 \end{aligned}$$

For more information about normal–half normal, normal-exponential, and normal-truncated models, see Kumbhakar and Lovell (2000); Coelli, Prasada Rao, and Battese (1998).

Heteroscedasticity and Box-Cox Transformation

Heteroscedasticity

If the variance of regression disturbance, (ϵ_i), is heteroscedastic, the variance can be specified as a function of variables

$$E(\epsilon_i^2) = \sigma_i^2 = f(\mathbf{z}_i'\boldsymbol{\gamma})$$

The following table shows various functional forms of heteroscedasticity and the corresponding options to request each model:

No.	Model	Options
1	$f(\mathbf{z}_i'\boldsymbol{\gamma}) = \sigma^2(1 + \exp(\mathbf{z}_i'\gamma))$	LINK=EXP (default)
2	$f(\mathbf{z}_i'\boldsymbol{\gamma}) = \sigma^2 \exp(\mathbf{z}_i'\gamma)$	LINK=EXP NOCONST
3	$f(\mathbf{z}_i'\boldsymbol{\gamma}) = \sigma^2(1 + \sum_{l=1}^{L} \gamma_l z_{li})$	LINK=LINEAR
4	$f(\mathbf{z}_i'\boldsymbol{\gamma}) = \sigma^2(1 + (\sum_{l=1}^{L} \gamma_l z_{li})^2)$	LINK=LINEAR SQUARE

For discrete choice models, σ^2 is normalized ($\sigma^2 = 1$) since this parameter is not identified. Note that in models 3 and 5, it may be possible that variances of some observations are negative. Although the QLIM procedure assigns a large penalty to move the optimization away from such region, it is possible that the optimization cannot improve the objective function value and gets locked in the region. Signs of such outcome include extremely small likelihood values or missing standard errors in the estimates. In models 2 and 6, variances are guaranteed to be greater or equal to zero, but it may be possible that variances of some observations are very close to zero. In these scenarios, standard errors may be missing. Models 1 and 4 do not have such problems. Variances in these models are always positive and never close to zero.

The heteroscedastic regression model is estimated using the log-likelihood function

$$\ell = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{N}\frac{1}{2}\ln(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{N}(\frac{e_i}{\sigma_i})^2$$

where $e_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}$.

Box-Cox Modeling

The Box-Cox transformation on x is defined as

$$x^{(\lambda)} = \begin{cases} \frac{x^{\lambda}-1}{\lambda} & \text{if} \lambda \neq 0 \\ \ln(x) & \text{if} \lambda = 0 \end{cases}$$

The Box-Cox regression model with heteroscedasticity is written as

$$\begin{aligned} y_i^{(\lambda_0)} &= \beta_0 + \sum_{k=1}^{K}\beta_k x_{ki}^{(\lambda_k)} + \epsilon_i \\ &= \mu_i + \epsilon_i \end{aligned}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$ and transformed variables must be positive. In practice, too many transformation parameters cause numerical problems in model fitting. It is common to have the same Box-Cox transformation performed on all the variables—that is, $\lambda_0 = \lambda_1 = \cdots = \lambda_K$. It is required for the magnitude of transformed variables to be in the tolerable range if the corresponding transformation parameters are $|\lambda| > 1$.

The log-likelihood function of the Box-Cox regression model is written as

$$\ell = -\frac{N}{2}\ln(2\pi) - \sum_{i=1}^{N}\ln(\sigma_i) - \frac{1}{2\sigma_i^2}\sum_{i=1}^{N}e_i^2 + (\lambda_0 - 1)\sum_{i=1}^{N}\ln(y_i)$$

where $e_i = y_i^{(\lambda_0)} - \mu_i$.

When the dependent variable is discrete, censored, or truncated, the Box-Cox transformation can be applied only to explanatory variables.

Bivariate Censored Dependent Variable Modeling

The generic form of a bivariate censored dependent variable model is

$$\begin{aligned} y_{1i}^* &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \epsilon_{1i} \\ y_{2i}^* &= \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \epsilon_{2i} \end{aligned}$$

where the disturbances, ϵ_{1i} and ϵ_{2i}, have a joint normal distribution with zero mean, standard deviations σ_1 and σ_2, and correlation ρ. y_1^* and y_2^* are latent variables. The dependent variables y_1 and y_2 might or might not be censored at the edges of the bivariate interval $\{[L_1, R_1], [L_2, R_2]\}$, depending on the behavior of the latent variables y_1^* and y_2^*:

$$y_{1i} = \begin{cases} R_1 & \text{if } R_1 < y_{1i}^* \\ y_{1i}^* & \text{if } L_1 \le y_{1i}^* \le R_1 \\ L_1 & \text{if } y_{1i}^* < L_1 \end{cases}$$

$$y_{2i} = \begin{cases} R_2 & \text{if } R_2 < y_{2i}^* \\ y_{2i}^* & \text{if } L_2 \le y_{2i}^* \le R_2 \\ L_2 & \text{if } y_{2i}^* < L_2 \end{cases}$$

There are three cases for the log likelihood of (y_{1i}, y_{2i}). The first case is where $y_{1i} = y_{1i}^*$ and $y_{2i} = y_{2i}^*$. That is, both observations are uncensored. The log likelihood is computed from a bivariate normal density,

$$\ell_i = \ln\left[\text{pdf}\left(y_{1i}^*, y_{2i}^*\right)\right] = \ln\left[\phi_2\left(\frac{y_{1i} - \mathbf{x_{1i}}'\boldsymbol{\beta}_1}{\sigma_1}, \frac{y_{2i} - \mathbf{x_{2i}}'\boldsymbol{\beta}_2}{\sigma_2}, \rho\right)\right] - \ln\sigma_1 - \ln\sigma_2$$

where $\phi_2(u, v, \rho)$ is the density function for a standardized bivariate normal distribution with correlation ρ,

$$\phi_2(u, v, \rho) = \frac{e^{-(1/2)(u^2+v^2-2\rho uv)/(1-\rho^2)}}{2\pi(1-\rho^2)^{1/2}}$$

The second case is where one variable is censored and one is not. For example, if $y_{1i} = y_{1i}^*$ and $y_{2i} = L_2$, then the log likelihood is computed as

$$\begin{aligned} \ell_i &= \ln\left[\int_{-\infty}^{L_2} \text{pdf}(y_{1i}^*, y_{2i}^*)dy_{2i}^*\right] = \ln\left[\int_{-\infty}^{L_2} \text{pdf}(y_{2i}^*|y_{1i}^*)\text{pdf}(y_{1i}^*)dy_{2i}^*\right] \\ &= \ln\left[\phi\left(\frac{y_{1i}^* - \mathbf{x_1}'\boldsymbol{\beta}_1}{\sigma_1}\right)\right] - \ln\sigma_1 + \ln\left[\Phi\left(\frac{L_2 - \mathbf{x_2}'\boldsymbol{\beta}_2 - \sigma_2\rho\frac{y_{1i}^* - \mathbf{x_1}'\boldsymbol{\beta}_1}{\sigma_1}}{\sigma_2\sqrt{1-\rho^2}}\right)\right] \end{aligned}$$

where ϕ and Φ are the density function and the cumulative probability function for a standardized univariate normal distribution, respectively.

The third case is where both dependent variables are censored. For example, if $y_{1i} = R_1$ and $y_{2i} = L_2$, then the log likelihood is

$$\ell_i = \ln \left[\int_{u=\frac{R_1 - \mathbf{x_1}'\boldsymbol{\beta}_1}{\sigma_1}}^{\infty} \int_{v=-\infty}^{\frac{L_2 - \mathbf{x_2}'\boldsymbol{\beta}_2}{\sigma_2}} \phi_2(u, v, \rho)\, du\, dv \right]$$

Selection Models

In sample selection models, one or several dependent variables are observed when another variable takes certain values. For example, the standard Heckman selection model can be defined as

$$z_i^* = \mathbf{w}_i'\boldsymbol{\gamma} + u_i$$

$$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \le 0 \end{cases}$$

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \quad \text{if } z_i = 1$$

where u_i and ϵ_i are jointly normal with 0 mean, standard deviations of 1 and σ, respectively, and correlation of ρ. Selection is based on the variable z, and y is observed when z has a value of 1. Least squares regression that uses the observed data of y produces inconsistent estimates of $\boldsymbol{\beta}$. The maximum likelihood method is used to estimate selection models. It is also possible to estimate these models by using Heckman's method, which is more computationally efficient. But it can be shown that the resulting estimates, although consistent, are not asymptotically efficient under a normality assumption. Moreover, this method often violates the constraint on the correlation coefficient $|\rho| \le 1$.

The log-likelihood function of the Heckman selection model is written as

$$\begin{aligned} \ell &= \sum_{i \in \{z_i = 0\}} \ln[1 - \Phi(\mathbf{w}_i'\boldsymbol{\gamma})] \\ &+ \sum_{i \in \{z_i = 1\}} \left\{ \ln \phi\left(\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma}\right) - \ln \sigma + \ln \Phi\left(\frac{\mathbf{w}_i'\boldsymbol{\gamma} + \rho \frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma}}{\sqrt{1 - \rho^2}} \right) \right\} \end{aligned}$$

The selection can be based on only one variable, but the selection can lead to several variables. For example, selection is based on the variable z in the following switching regression model:

$$z_i^* = \mathbf{w}_i'\boldsymbol{\gamma} + u_i$$

$$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \le 0 \end{cases}$$

$$\begin{aligned} y_{1i} &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \epsilon_{1i} \quad \text{if } z_i = 0 \\ y_{2i} &= \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \epsilon_{2i} \quad \text{if } z_i = 1 \end{aligned}$$

If $z = 0$, then y_1 is observed. If $z = 1$, then y_2 is observed. Because $y1$ and $y2$ are never observed at the same time, the correlation between y_1 and y_2 cannot be estimated. Only the correlation between z and y_1 and the correlation between z and y_2 can be estimated. This estimation uses the maximum likelihood method.

A brief example of the SAS statements for this model can be found in "Example 28.4: Sample Selection Model" on page 2047.

The Heckman selection model can include censoring or truncation. For a brief example of the SAS statements for these models, see "Example 28.5: Sample Selection Model with Truncation and Censoring" on page 2048. The following example shows a variable y_i that is censored from below at zero:

$$z_i^* = \mathbf{w}_i'\boldsymbol{\gamma} + u_i$$

$$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \le 0 \end{cases}$$

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \quad \text{if } z_i = 1$$

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \le 0 \end{cases}$$

In this case, the log-likelihood function of the Heckman selection model needs to be modified as follows to include the censored region:

$$\begin{aligned} \ell \;=\; & \sum_{\{i\,|\,z_i=0\}} \ln[1-\Phi(\mathbf{w}_i'\boldsymbol{\gamma})] \\ & + \sum_{\{i\,|\,z_i=1, y_i=y_i^*\}} \left\{ \ln\left[\phi\left(\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma}\right)\right] - \ln\sigma + \ln\left[\Phi\left(\frac{\mathbf{w}_i'\boldsymbol{\gamma} + \rho\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma}}{\sqrt{1-\rho^2}}\right)\right]\right\} \\ & + \sum_{\{i\,|\,z_i=1, y_i=0\}} \ln \int_{-\infty}^{\frac{-\mathbf{x_i}'\boldsymbol{\beta}}{\sigma}} \int_{-\mathbf{w_i}'\gamma}^{\infty} \phi_2(u,v,\rho)\,du\,dv \end{aligned}$$

In case y_i is truncated from below at 0 instead of censored, the likelihood function can be written as

$$\begin{aligned} \ell \;=\; & \sum_{\{i\,|\,z_i=0\}} \ln[1-\Phi(\mathbf{w}_i'\boldsymbol{\gamma})] \\ & + \sum_{\{i\,|\,z_i=1\}} \left\{ \ln\left[\phi\left(\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma}\right)\right] - \ln\sigma + \ln\left[\Phi\left(\frac{\mathbf{w}_i'\boldsymbol{\gamma} + \rho\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma}}{\sqrt{1-\rho^2}}\right)\right] - \ln\left[\Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)\right]\right\} \end{aligned}$$

Heckman's Two-Step Selection Method

Sample selection bias arises from nonrandom selection of the sample from the population. A classic example is using a sample of market wages for working women to estimate female labor supply function. This sample is nonrandom because it includes only the wages of women whose market wage exceeds their home wage at zero hours of work.

A simple selection model can be written as the latent model

$$z_i^* = \mathbf{w}_i'\boldsymbol{\gamma} + u_i$$

$$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases}$$

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \quad \text{if } z_i = 1$$

where u_i and ϵ_i are jointly normal with 0 mean, standard deviations of 1 and σ, respectively, and correlation of ρ. The dependent variable y_i (wage) is observed if the latent variable z_i^* (the difference between market wage and reservation wage) is positive or if the indicator variable z_i (labor force participation) is 1.

The model of interest that applies to the observations in the selected sample can be written as

$$E(y_i|\mathbf{x}_i, z_i = 1) = \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma\lambda(\mathbf{w}_i'\boldsymbol{\gamma})$$

where $\lambda(\mathbf{w}_i'\boldsymbol{\gamma}) = \phi(\mathbf{w}_i'\boldsymbol{\gamma})/\Phi(\mathbf{w}_i'\boldsymbol{\gamma})$. Hence, the following regression equation is valid for the observations for which $z_i = 1$:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma\lambda(\mathbf{w}_i'\boldsymbol{\gamma}) + v_i$$

Therefore, estimates of $\boldsymbol{\beta}$ that are obtained from the OLS regression of y on $\mathbf{x}$ by using the selected sample (that is, the sample for which $z_i = 1$) suffer from omitted variable bias if selection bias is really the case. Although maximum likelihood estimation of $\boldsymbol{\beta}$ is consistent and efficient, Heckman's two-step method is more frequently used. Heckman's two-step method can be requested by specifying the HECKIT option of the QLIM statement.

Heckman's two-step method is as follows:

1. Obtain $\hat{\boldsymbol{\gamma}}$, the estimate of the parameters of the probability that $z_i^* > 0$, by using regressors $\mathbf{w}_i$ and the binary dependent variable z_i by probit analysis for the full sample. Compute $\hat{\lambda}_i = \lambda(\mathbf{w}_i'\hat{\boldsymbol{\gamma}})$.
2. Obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\beta}_\lambda$, the estimates of $\boldsymbol{\beta}$ and $\rho\sigma$, by least squares regression of y_i on $\mathbf{x}_i$ and $\hat{\lambda}_i$ by using observations on the selected subsample.

The standard least squares estimators of the population variance σ^2 and the variances of the estimated coefficients are incorrect. To test hypotheses, the correct ones need to be calculated. An estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{N_1}\sum_{i=1}^{N_1} e_i^2 + \hat{\beta}_\lambda^2 \frac{1}{N_1}\sum_{i=1}^{N_1} \hat{\delta}_i$$

where N_1 is the selected subsample size, e_i is the residual for the ith observation obtained from step 2, and $\hat{\delta}_i = \hat{\lambda}_i^2 + \hat{\lambda}_i \mathbf{w}_i'\hat{\boldsymbol{\gamma}}$. Let $\mathbf{X}_*$ be an $N_1 \times (K+1)$ matrix with ith row $[\mathbf{x}_i' \;\; \lambda_i]$, and define $\mathbf{W}$ similarly with ith row $\mathbf{w}_i'$. Then the estimator of the asymptotic covariance of $[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_\lambda]$ is

$$\text{EstAsyVar}[\hat{\boldsymbol{\beta}}, \hat{\beta}_\lambda] = \hat{\sigma}^2[\mathbf{X}_*'\mathbf{X}_*]^{-1}[\mathbf{X}_*'(\mathbf{I} - \hat{\rho}^2\hat{\boldsymbol{\Delta}})\mathbf{X}_* + \mathbf{Q}][\mathbf{X}_*'\mathbf{X}_*]^{-1}$$

where $\hat{\rho}^2 = \hat{\beta}_\lambda^2/\hat{\sigma}^2$, $\hat{\boldsymbol{\Delta}} = \text{diag}(\hat{\delta}_i)$, and

$$\mathbf{Q} = \hat{\sigma}^2(\mathbf{X}_*'\hat{\boldsymbol{\Delta}}\mathbf{W})\text{Est.Asy.Var}(\hat{\boldsymbol{\gamma}})(\mathbf{W}'\hat{\boldsymbol{\Delta}}\mathbf{X}_*)$$

where Est.Asy.Var($\hat{\boldsymbol{\gamma}}$) is the estimator of the asymptotic covariance of the probit coefficients that are obtained in step 1. When you specify the HECKIT option, PROC QLIM uses a numerical estimated asymptotic variance.

When the HECKIT option is specified, PROC QLIM reports the corrected standard errors for $[\hat{\boldsymbol{\beta}}, \hat{\beta}_\lambda]$ automatically. However, if you need the conventional OLS standard errors, you can specify the HECKIT(UNCORRECTED) option.

In the selected regression model, when the coefficient of $\lambda(\mathbf{w}_i'\boldsymbol{\gamma})$ is 0, you do not need Heckman's two-step estimation method; a simple regression of y on $\mathbf{x}$ produces consistent estimates for $\boldsymbol{\beta}$, and the OLS standard errors are correct. Thus, a standard t test on $\hat{\beta}_\lambda$ (which uses the estimate from step 2 and the uncorrected standard errors) is a valid test of the null hypothesis of no selection bias.

Although Heckman's two-step method uses the OLS method in the second stage, you can request the ML method by specifying the HECKIT(SECONDSTAGE=ML) option. When the second-stage method is the ML method, the model for y_i can be nonlinear.

Multivariate Limited Dependent Models

The multivariate model is similar to bivariate models. The generic form of the multivariate limited dependent variable model is

$$\begin{aligned} y_{1i}^* &= \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \epsilon_{1i} \\ y_{2i}^* &= \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \epsilon_{2i} \\ &\dots \\ y_{mi}^* &= \mathbf{x}_{mi}'\boldsymbol{\beta}_m + \epsilon_{mi} \end{aligned}$$

where m is the number of models to be estimated. The vector ϵ has multivariate normal distribution with mean 0 and variance-covariance matrix Σ. Similar to bivariate models, the likelihood may involve computing multivariate normal integrations. This is done using Monte Carlo integration. (See Genz 1992; Hajivassiliou and McFadden 1998.)

When the number of equations, N, increases in a system, the number of parameters increases at the rate of N^2 because of the correlation matrix. When the number of parameters is large, sometimes the optimization converges but some of the standard deviations are missing. This usually means that the model is over-parameterized. The default method for computing the covariance is to use the inverse Hessian matrix. The Hessian is computed by finite differences, and in over-parameterized cases, the inverse cannot be computed. It is recommended that you reduce the number of parameters in such cases. Sometimes using the outer product covariance matrix (COVEST=OP option) might also help.

Variable Selection

Variable Selection

Variable Selection

Variable selection uses either Akaike's information criterion (AIC) or the Schwartz Bayesian criterion (SBC) and either a forward selection method or a backward elimination method.

Forward selection starts from a small subset of variables. In each step, the variable that gives the largest decrease in the value of the information criterion specified in the CRITER= option (AIC or SBC) is added. The process stops when the next candidate to be added does not reduce the value of the information criterion by more than the amount specified in the LSTOP= option in the MODEL statement.

Backward elimination starts from a larger subset of variables. In each step, one variable is dropped based on the information criterion that is chosen.

Tests on Parameters

Tests on Parameters

Tests on Parameters

In general, the hypothesis tested can be written as

$$H_0 : \mathbf{h}(\theta) = 0$$

where $\mathbf{h}(\theta)$ is an r by 1 vector valued function of the parameters θ given by the r expressions specified in the TEST statement.

Let $\hat{V}$ be the estimate of the covariance matrix of $\hat{\theta}$. Let $\hat{\theta}$ be the unconstrained estimate of θ and $\tilde{\theta}$ be the constrained estimate of θ such that $h(\tilde{\theta}) = 0$. Let

$$A(\theta) = \partial h(\theta)/\partial\theta \mid_{\hat{\theta}}$$

Using this notation, the test statistics for the three kinds of tests are computed as follows.

The Wald test statistic is defined as

$$W = h'(\hat{\theta}) \left(A(\hat{\theta})\hat{V}A'(\hat{\theta}) \right)^{-1} h(\hat{\theta})$$

The Wald test is not invariant to reparameterization of the model (Gregory and Veall 1985, Gallant 1987, p. 219). For more information about the theoretical properties of the Wald test, see Phillips and Park (1988).

The Lagrange multiplier test statistic is

$$LM = \lambda' A(\tilde{\theta})\tilde{V}A'(\tilde{\theta})\lambda$$

where λ is the vector of Lagrange multipliers from the computation of the restricted estimate $\tilde{\theta}$.

The likelihood ratio test statistic is

$$LR = 2\left(L(\hat{\theta}) - L(\tilde{\theta})\right)$$

where $\tilde{\theta}$ represents the constrained estimate of θ and L is the concentrated log-likelihood value.

For each kind of test, under the null hypothesis the test statistic is asymptotically distributed as a χ^2 random variable with r degrees of freedom, where r is the number of expressions in the TEST statement. The p-values reported for the tests are computed from the $\chi^2(r)$ distribution and are only asymptotically valid.

Monte Carlo simulations suggest that the asymptotic distribution of the Wald test is a poorer approximation to its small sample distribution than that of the other two tests. However, the Wald test has the lowest computational cost, since it does not require computation of the constrained estimate $\tilde{\theta}$.

The following is an example of using the TEST statement to perform a likelihood ratio test:

```
proc qlim;
   model y = x1 x2 x3;
   test x1 = 0, x2 * .5 + 2 * x3 = 0 /lr;
run;
```

Endogeneity and Instrumental Variables

The PROC QLIM models such as qualitative response or limited dependent variable models assume that the errors are independent of the explanatory variables. If this assumption fails to hold, the distributional form that the likelihood is based on is misspecified and the obtained coefficients are inconsistent.

To begin, consider a linear model

$$y_i = y_i^* = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

Assume that $E(u) = 0$, $\mathrm{Cov}(x_j, u) = 0$ for $j = 1, \ldots, k-1$, and $\mathrm{Cov}(x_k, u) = \rho \neq 0$. Therefore, x_k is endogenous. The endogeneity comes from many sources, such as x_k having measurement error or omitting a variable that is correlated with x_k. If you ignore the endogeneity, you can estimate this model in PROC QLIM as follows (assuming $k = 4$):

```
proc qlim data=a;
   model y = x1 x2 x3 x4;
run;
```

However, this approach produces inconsistent maximum likelihood estimates. To obtain consistent maximum likelihood estimates, you should consider the joint density of the dependent variable and the endogenous variables. To do this in PROC QLIM, you need at least one instrument—that is, an observable variable, z_1—that is not in the structural equation and that satisfies two conditions: z_1 is exogenous (that is, $\mathrm{Cov}(z_1, u) = 0$), and z_1 must be correlated with the endogenous regressor x_k. Then, you can model x_k as

$$x_{ki} = \pi_0 + \pi_1 x_{1i} + \cdots + \pi_{k-1} x_{(k-1)i} + \theta z_{1i} + \epsilon_i$$

You can now write this reduced form equation along with the structural equation to obtain the consistent maximum likelihood estimates as follows:

```
proc qlim data=a;
   model y = x1 x2 x3 x4;
   model x4 = x1 x2 x3 z1;
run;
```

Estimating the structural model together with the reduced form models for the endogenous explanatory variables gives you the full information maximum likelihood (FIML) estimates. Because of the linearity of

the structural model, you can estimate it efficiently and more simply by using the two-stage least squares estimator. However, PROC QLIM handles nonlinear models such as qualitative response and limited dependent variable models, and in their estimation it maximizes the corresponding joint likelihood function (for more information and an application, see Wooldridge 2010, Section 15.7.3). In the case of endogeneity, when the reduced form models for the endogenous explanatory variables are written along with the structural model, PROC QLIM maximizes the likelihood function that is obtained from the joint density of the response variable and the endogenous explanatory variables. For example, consider the following censored regression model in which one of the explanatory variables is a continuous endogenous variable:

$$\begin{aligned} y_{1i}^* &= \alpha y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta} + u_i \\ y_{2i} &= \mathbf{z}_i'\boldsymbol{\pi} + \epsilon_i \\ y_{1i} &= \begin{cases} y_{1i}^* & \text{if } y_{1i}^* > 0 \\ 0 & \text{if } y_{1i}^* \le 0 \end{cases} \end{aligned}$$

The exogenous explanatory variables are $\mathbf{z}_{1i}$, and the continuous endogenous explanatory variable is y_{2i}.

The likelihood function to maximize is

$$L = \prod_{i \in \{y_{1i} > 0\}} f(y_{1i}, y_{2i}) \cdot \prod_{i \in \{y_{1i} = 0\}} \int_{-\infty}^{0} f(y_{1i}^*, y_{2i}) dy_{1i}^*$$

where $f(y_{1i}^*, y_{2i})$ is the joint density of y_{1i}^* and y_{2i}. Note that y_{1i} is substituted for y_{1i}^* when $y_{1i} > 0$. If you assume $(u_i, \epsilon_i) \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_u^2 & \eta \\ \eta & \sigma_\epsilon^2 \end{bmatrix}$, then, by using $f(y_{1i}^*, y_{2i}) = f(y_{1i}^* | y_{2i}) \cdot f(y_{2i})$, you can write the likelihood function for each i as a multiplication of two parts. The first part is the probability density function of the normal distribution with mean $\mathbf{z}_i'\boldsymbol{\pi}$ and variance σ_ϵ^2, and the second part follows a Tobit model that has latent mean $\alpha y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\pi} + (\eta/\sigma_\epsilon^2)(y_{2i} - \mathbf{z}_i'\boldsymbol{\pi})$ and variance $\sigma_u^2 - (\eta^2/\sigma_\epsilon^2)$. Then, you can obtain the log-likelihood function by taking the log of this multiplication and summing over i (for more information, see Wooldridge 2002, Section 16.6.2). This is the log-likelihood function that PROC QLIM maximizes. The parameters $(\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}}, \hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2, \hat{\eta})$ that are obtained from this maximization are the FIML estimators. Assuming that the latent model includes two instrumental variables and two exogenous explanatory variables, you can estimate this model in PROC QLIM as follows:

```
proc qlim data=a;
   model y1 = y2 z11 z12 / censored(lb=0);
   model y2 = z11 z12 z21 z22;
run;
```

For simple examples like the preceding ones, you can derive the likelihood function easily. However, as the number of endogenous explanatory variables increases, if these variables have a discontinuous nature, if simultaneity among equations exists, or if a combination of these occurs, then the derivation of the likelihood function becomes cumbersome, or, in some cases, the likelihood function does not even have a closed analytical form.

PROC QLIM can handle endogeneity regardless of the nature of the endogenous explanatory variables for a single structural model. In the case of one endogenous explanatory variable, PROC QLIM reports the FIML estimates that are calculated by using the analytical likelihood function that is obtained from the joint distribution of the dependent variable and the endogenous variable. When there is more than one endogenous explanatory variable, the analytical form of the likelihood function is usually not available; in this case PROC QLIM reports the simulated maximum likelihood estimates. For the simulated maximum likelihood

estimation method, PROC QLIM uses the Geweke-Hajivassiliou-Keane (GHK) simulator (see, among others, Hajivassiliou, McFadden, and Ruud 1996) to simulate the joint distribution of the dependent variable and the endogenous variables. The simulation is facilitated by assuming that the error terms in the latent models for the dependent variable and the endogenous explanatory variables are distributed as multivariate normal.

When you estimate a model in PROC QLIM, you can take the endogeneity into account by writing the structural model along with the reduced form models for each endogenous variable. Examples are provided in the following sections.

Probit Model with a Continuous Endogenous Explanatory Variable

Consider a probit model that contains a single endogenous explanatory variable in addition to two instruments and two exogenous explanatory variables. The model is

$$\begin{aligned} y^*_{1i} &= \alpha_1 y_{2i} + \beta_1 z_{1i} + \beta_2 z_{2i} + u_i \\ y^*_{2i} &= \pi_1 z_{1i} + \pi_2 z_{2i} + \pi_3 z_{3i} + \pi_4 z_{4i} + \epsilon_i \\ y_{1i} &= \begin{cases} 1 & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \le 0 \end{cases} \\ y_{2i} &= y^*_{2i} \end{aligned}$$

where $\mathrm{Cov}(u, \epsilon) = \eta$. You can estimate this model by using the following statements:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / discrete;
   model y2 = z1 z2 z3 z4;
run;
```

Probit Model with a Binary Endogenous Explanatory Variable

Consider a probit model that contains a single binary endogenous explanatory variable in addition to two instruments and two exogenous explanatory variables. The model is

$$\begin{aligned} y^*_{1i} &= \alpha_1 y_{2i} + \beta_1 z_{1i} + \beta_2 z_{2i} + u_i \\ y^*_{2i} &= \pi_1 z_{1i} + \pi_2 z_{2i} + \pi_3 z_{3i} + \pi_4 z_{4i} + \epsilon_i \\ y_{1i} &= \begin{cases} 1 & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \le 0 \end{cases} \\ y_{2i} &= \begin{cases} 1 & \text{if } y^*_{2i} > 0 \\ 0 & \text{if } y^*_{2i} \le 0 \end{cases} \end{aligned}$$

where $\mathrm{Cov}(u, \epsilon) = \eta$. You can estimate this model by using the following statements:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / discrete;
   model y2 = z1 z2 z3 z4 / discrete;
run;
```

Probit Model with a Censored Endogenous Explanatory Variable

Consider a probit model that contains a single censored (below zero) endogenous explanatory variable in addition to two instruments and two exogenous explanatory variables. The model is

$$
\begin{aligned}
y^*_{1i} &= \alpha_1 y_{2i} + \beta_1 z_{1i} + \beta_2 z_{2i} + u_i \\
y^*_{2i} &= \pi_1 z_{1i} + \pi_2 z_{2i} + \pi_3 z_{3i} + \pi_4 z_{4i} + \epsilon_i \\
y_{1i} &= \begin{cases} 1 & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \le 0 \end{cases} \\
y_{2i} &= \begin{cases} y^*_{2i} & \text{if} y^*_{2i} > 0 \\ 0 & \text{if} y^*_{2i} \le 0 \end{cases}
\end{aligned}
$$

where $\mathrm{Cov}(u, \epsilon) = \eta$. You can estimate this model by using the following statements:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / discrete;
   model y2 = z1 z2 z3 z4 / censored(lb=0);
run;
```

Censored Regression Model with a Binary Endogenous Explanatory Variable

Consider a Type 1 Tobit model that contains a single binary endogenous explanatory variable in addition to two instruments and two exogenous explanatory variables. The model is

$$
\begin{aligned}
y^*_{1i} &= \alpha_1 y_{2i} + \beta_1 z_{1i} + \beta_2 z_{2i} + u_i \\
y^*_{2i} &= \pi_1 z_{1i} + \pi_2 z_{2i} + \pi_3 z_{3i} + \pi_4 z_{4i} + \epsilon_i \\
y_{1i} &= \begin{cases} y^*_{1i} & \text{if} y^*_{1i} > 0 \\ 0 & \text{if} y^*_{1i} \le 0 \end{cases} \\
y_{2i} &= \begin{cases} 1 & \text{if } y^*_{2i} > 0 \\ 0 & \text{if } y^*_{2i} \le 0 \end{cases}
\end{aligned}
$$

where $\mathrm{Cov}(u, \epsilon) = \eta$. You can estimate this model by using the following statements:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / censored(lb=0);
   model y2 = z1 z2 z3 z4 / discrete;
run;
```

Censored Regression Model with Binary and Continuous Endogenous Explanatory Variables

Consider a Type 1 Tobit model that contain binary and continuous endogenous explanatory variables in addition to two instruments and two exogenous explanatory variables. The model is

$$
\begin{aligned}
y^*_{1i} &= \alpha_1 y_{21i} + \alpha_2 y_{22i} + \beta_1 z_{1i} + \beta_2 z_{2i} + u_i \\
y^*_{21i} &= \pi_{11} z_{1i} + \pi_{12} z_{2i} + \pi_{13} z_{3i} + \pi_{14} z_{4i} + \epsilon_{1i} \\
y^*_{22i} &= \pi_{21} z_{1i} + \pi_{22} z_{2i} + \pi_{23} z_{3i} + \pi_{24} z_{4i} + \epsilon_{2i} \\
y_{1i} &= \begin{cases} y^*_{1i} & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \le 0 \end{cases} \\
y_{21i} &= \begin{cases} 1 & \text{if } y^*_{21i} > 0 \\ 0 & \text{if } y^*_{21i} \le 0 \end{cases} \\
y_{22i} &= y^*_{22i}
\end{aligned}
$$

where $\mathrm{Cov}(u, \epsilon_1, \epsilon_2) = \eta$. You can estimate this model by using the following statements:

```
proc qlim data=a;
   model y1 = y21 y22 z1 z2 / censored(lb=0);
   model y21 = z1 z2 z3 z4  / discrete;
   model y22 = z1 z2 z3 z4;
run;
```

Probit Model with Binary, Censored, and Truncated Endogenous Explanatory Variables

Consider a probit model that contains binary, censored (below zero), and truncated (below zero) endogenous explanatory variables. The model is

$$
\begin{aligned}
y^*_{1i} &= \alpha_1 y_{21i} + \alpha_2 y_{22i} + \alpha_3 y_{23i} + u_i \\
y^*_{21i} &= \pi_{11} z_{1i} + \pi_{12} z_{2i} + \pi_{13} z_{3i} + \pi_{14} z_{4i} + \epsilon_{1i} \\
y^*_{22i} &= \pi_{21} z_{1i} + \pi_{22} z_{2i} + \pi_{23} z_{3i} + \pi_{24} z_{4i} + \epsilon_{2i} \\
y^*_{23i} &= \pi_{31} z_{1i} + \pi_{32} z_{2i} + \pi_{33} z_{3i} + \pi_{34} z_{4i} + \epsilon_{3i} \\
y_{1i} &= \begin{cases} 1 & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \le 0 \end{cases} \\
y_{21i} &= \begin{cases} 1 & \text{if } y^*_{21i} > 0 \\ 0 & \text{if } y^*_{21i} \le 0 \end{cases} \\
y_{22i} &= \begin{cases} y^*_{22i} & \text{if} y^*_{22i} > 0 \\ 0 & \text{if } y^*_{22i} \le 0 \end{cases} \\
y_{23i} &= y^*_{23i} \quad \text{if} \quad y^*_{23i} > 0
\end{aligned}
$$

where $z_1, \ldots, z_4$ are the instrumental variables that are independent of the errors. You can estimate this model by using the following statements:

```
proc qlim data=a;
   model y1  = y21 y22 y23 / discrete;
   model y21 = z1 z2 z3 z4 / discrete;
   model y22 = z1 z2 z3 z4 / censored(lb=0);
   model y23 = z1 z2 z3 z4 / truncated(lb=0);
run;
```

Note that the dependent variable y_1 should not occur in the models for the endogenous explanatory variables, because this causes inconsistent coefficient estimates. In other words, you should write the models for the endogenous explanatory variables as reduced form models. PROC QLIM does not handle simultaneous equations models.

Test for Endogeneity

PROC QLIM has two ways to test the null hypothesis that an endogenous explanatory variable (EEV) is in fact exogenous. In the case of a single EEV, the first testing method involves a likelihood ratio test of H_0 : _rho $= 0$. For example, consider the probit model with a binary endogenous explanatory variable that was considered earlier; y_2 is exogenous if the error term in the model for y_1^* is uncorrelated with the error term in the model for y_2^*. Therefore, testing to determine whether this correlation is 0 or not provides an endogeneity test for y_2. You can do this in PROC QLIM as follows:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / discrete;
   model y2 = z1 z2 z3 z4 / discrete;
   test  _rho = 0 / LR;
run;
```

Failing to reject the null hypothesis favors the decision that $y2$ is exogenous in the model for $y1$.

When there are two or more EEVs, the test becomes the joint likelihood ratio test of whether corresponding correlations are 0 or not.

The second testing method is similar to the approach of Rivers and Vuong (1988). Considering the same model, you can write

$$u_i = \theta\epsilon_i + e_i$$

where $\theta = \eta/\sigma_\epsilon^2$ and e is independent of zs and ϵ. You can now write

$$y_{1i}^* = \alpha_1 y_{2i} + \beta_1 z_{1i} + \beta_2 z_{2i} + \theta\epsilon_i + e_i$$

Testing $H_0 : \theta = 0$ is the same as testing whether u_i is correlated with ϵ_i or testing whether y_{2i} is endogenous or not. Because ϵ_i are unobserved, you can replace them with the OLS residuals from the model for y_{2i}^* and apply a robust t test. Note that even though y_{2i} is binary (or censored), the test is still correct under H_0.

This approach can be summarized as a two-step procedure. In the first step, generated regressors—that is, the OLS residuals from the models for each of the EEVs—are obtained. In the second step, the structural model that includes the generated regressors as additional explanatory variables is estimated by the maximum likelihood method and the joint significance of these generated regressors is tested by the Wald test.

In PROC QLIM, you can apply the second method for the same test that was considered previously as follows:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / discrete endotest(y2);
   model y2 = z1 z2 z3 z4 / discrete;
run;
```

Overidentification Test

In PROC QLIM you can test the validity of instrumental variables (IVs) by specifying the OVERID option in the ENDOGENOUS or MODEL statement. The OVERID test is a maximum likelihood version of the overidentifying restrictions test in the IV framework. If you have more IVs than are necessary for identification—that is, overidentifying IVs—you can use them to test the validity of your IVs. When you use the OVERID option to specify the overidentifying IVs, it applies the likelihood ratio test of the joint significance of these IVs, included as additional explanatory variables in the structural model that it estimates by the MLE jointly with the reduced form models. In effect, you test whether the overidentifying IVs are correlated with the error term in the structural model. You specify the reduced form models through the overidentifying IVs. The structural model is the model that includes the OVERID option. For example, consider the probit model that contains a continuous endogenous explanatory variable. You can consider $z3$ or $z4$ in the model for $y2$ as an overidentifying IV; therefore, you can specify the OVERID test as follows:

```
proc qlim data=a;
   model y1 = y2 z1 z2 / discrete overid(y2.z4);
   model y2 = z1 z2 z3 z4;
run;
```

In this case, PROC QLIM estimates the structural model $y1$, including the overidentifying IV $z4$ as an additional explanatory variable in this model, jointly with the reduced form model $y2$. Then it uses the likelihood ratio test to test the hypothesis that the overidentifying IV is insignificant. Rejecting this hypothesis raises doubts about the validity of the instruments $z3$ and $z4$.

Note that, as long as you have continuous endogenous explanatory variables, the test result is invariant to which overidentifying IVs you specify in the test.

Random-Parameters Models and Panel Data Analysis

Consider the effect of age on an individual's health self-assessment that is recorded using the values $0, 1, \ldots, 10$, where 0 indicates the poorest health. You can model the self-assessment outcome by an ordered probit or logit in PROC QLIM by using the option DISCRETE(D=NORMAL) or DISCRETE(D=LOGISTIC) in the MODEL or ENDOGENOUS statement.

One important shortcoming of this traditional way of modeling is the underlying assumption that, for all individuals, the explanatory variables have fixed constant coefficients. This assumption implies that the impact of the explanatory variables on the dependent variable is the same for all the individuals. However, the

assumption might not be realistic, because individuals are usually heterogeneous and hence the coefficient values are expected to vary across the individual observations. In the health self-assessment example, it is expected that aging involves cognitive and physical decline, so on average the relationship between age and health is expected to be negative. However, believing that this negative relationship is the same for every individual ignores the fact that for some individuals aging brings wiser life choices, including a healthier lifestyle and improved emotional well-being, and hence even improved health. Thus, enforcing a negative relationship can cause misleading inferences for this subgroup of individuals with a positive coefficient. Similarly, the effect might be negative for every individual, but its magnitude can vary across observations. In any case, if you are modeling such a behavior, then taking into account the unobserved heterogeneity, where parameter values vary across the observations because of unobserved factors, is more likely to give you more realistic results.

Random-parameters models accommodate such a heterogeneity by allowing the coefficients to vary randomly across individuals based on some prespecified distribution, $h(\boldsymbol{\theta})$. The set of parameters $\boldsymbol{\theta}$ defines the unobserved heterogeneity. Therefore, the goal is to estimate those parameters to define the individual heterogeneity.

If you have panel data, you can include random parameters by using the RANDOM statement for all the single-equation models of PROC QLIM—binary probit or logit, ordered probit or logit, Tobit (censored and truncated), stochastic frontier production and cost, and linear regression models—to generalize these models further in order to obtain more realistic results. However, you do not have to have the observations collected in a panel data setting to apply random-parameters models in PROC QLIM. The random-parameters models can also be applied in cross-sectional data as long as you specify the group or subject variable across which the parameter heterogeneity occurs.

General Models with Random Parameters

Random-parameters models allow individual heterogeneity in the coefficients in the latent process,

$$y^*_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}_i + v_{it}$$

where y^*_{it} is a latent variable, $\mathbf{x}_{it}$ is a vector of covariates, and v_{it} is the error term. In the applications for a panel data set, the subscript i represents individuals and t represents the time period.

The model assumes that parameters are randomly distributed with mean

$$E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$$

and variance

$$\mathrm{Var}(\boldsymbol{\beta}_i) = \boldsymbol{\Omega}$$

$\boldsymbol{\Omega}$ is a positive definite matrix. If the random parameters are not correlated with one another, then $\boldsymbol{\Omega}$ becomes a diagonal matrix. Let $\boldsymbol{\Gamma}$ be the Cholesky factorization of the covariance matrix of the random parameters, $\boldsymbol{\Omega} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$. In other words, $\boldsymbol{\Gamma}$ is the lower triangular matrix that produces $\boldsymbol{\Omega}$. By construction,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\boldsymbol{\omega}_i$$

where $\boldsymbol{\omega}_i$ is a random vector with zero means and unit standard deviations. In the no-correlation case, $\boldsymbol{\Gamma}$ is also a diagonal matrix with the standard deviations of $\boldsymbol{\omega}_i$ on the diagonal.

PROC QLIM assumes that $\boldsymbol{\omega}_i$ are normally distributed; hence $\boldsymbol{\beta}_i$ is normally distributed with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Omega}$.

Some of the explanatory variables in the latent model might have fixed (nonrandom) coefficients. In this case $\boldsymbol{\beta}_i$ can be written conveniently as

$$\boldsymbol{\beta}_i = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 + \boldsymbol{\Gamma}\boldsymbol{\omega}_i \end{pmatrix}$$

where $\boldsymbol{\beta}_1$ is the vector of nonrandom (fixed) coefficients and $\boldsymbol{\beta}_2$ is the vector of the means of the random coefficients.

The general form of the conditional density for the observed response can be written as

$$f(y_{it}|x_{it}, \boldsymbol{\omega}_i) = g(y_{it}, x_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is the parameter vector that includes the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$; the standard deviation of v_{it}, σ; and other parameters specified by the model.

The joint density for the ith group conditional on $\boldsymbol{\omega}$ and $\mathbf{x}_i$ is

$$f(y_{i1}, y_{i2}, \ldots, y_{iT_i}|\mathbf{x}_i, \boldsymbol{\omega}_i; \boldsymbol{\theta}) = \prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta})$$

Because $\boldsymbol{\omega}_i$ is unobserved, it is necessary to obtain the unconditional likelihood by taking the expectation of this likelihood over the distribution of $\boldsymbol{\omega}_i$. Thus

$$L_i = f(y_{i1}, y_{i2}, \ldots, y_{iT_i}|\mathbf{x}_i; \boldsymbol{\theta}) = \int_{\boldsymbol{\omega}} \left[\prod_{t=1}^{T_i} g(y_{it}, x_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta}) \right] h(\boldsymbol{\omega}_i; \boldsymbol{\theta}) d\boldsymbol{\omega}$$

where $h(\boldsymbol{\omega}_i; \boldsymbol{\theta})$ is the probability density function of $\boldsymbol{\omega}_i$. Under the normality assumption, $h(\boldsymbol{\omega}_i; \boldsymbol{\theta}) = \phi(\boldsymbol{\omega}_i)$, where $\phi(\cdot)$ is the probability density function of the standard normal distribution. The true log-likelihood function is obtained by summing $\ln L_i$, the log of the contribution of the ith individual to the total, over the individuals:

$$\ln L = \sum_{i=1}^{N} \ln L_i = \sum_{i=1}^{N} \ln \left[\int_{\boldsymbol{\omega}} \left(\prod_{t=1}^{T_i} g(y_{it}, x_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta}) \right) \phi(\boldsymbol{\omega}_i) d\boldsymbol{\omega} \right]$$

The integral in the square brackets does not have a closed form, so it is difficult to perform maximum likelihood estimation. However, this integration can be approximated and likelihood estimation is still possible. The subsection "Estimation" on page 2013 discusses various methods of approximation for this integral.

The nature of the dependent variable specifies the log-likelihood function. For example, if the dependent variable is binary and its probability is defined by a normal distribution (a probit model), then

$$g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta}) = \Phi[(2y_{it} - 1)(x'_{it}\boldsymbol{\beta}_i)]$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. If the dependent variable is modeled by a logit, then

$$g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta}) = \Lambda[(2y_{it} - 1)(x'_{it}\boldsymbol{\beta}_i)]$$

where $\Lambda(\cdot)$ is the cumulative density function of the standard logistic distribution.

The likelihood function is maximized by solving the likelihood equations

$$\frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{N} \frac{\partial \ln L_i}{\partial \boldsymbol{\theta}}$$

These derivatives involve integration. The integration is approximated by the same method that is used to calculate the likelihood.

When you use one of the simulation methods that are described in the subsections "Monte Carlo Integration" on page 2013 and "QMC Method Using the Halton Sequence" on page 2014, the log likelihood to be optimized becomes

$$\ln L_{\text{simulated}} = \sum_{i=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \left(\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta}) \right) \right]$$

The general formulation of the gradients is

$$\frac{\partial \ln L_{\text{simulated}}}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{N} \frac{\frac{1}{R} \sum_{r=1}^{R} \frac{\partial \prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{\frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\omega}_i; \boldsymbol{\theta})}$$

The formulation of the derivatives with respect to each type of parameter differs from model to model.

Note that $\boldsymbol{\theta}$ includes the elements of $\boldsymbol{\Gamma}$ rather than $\boldsymbol{\Omega}$. That is, the optimization is performed with respect to elements of $\boldsymbol{\Gamma}$. Therefore, when you use the ITPRINT option, the resulting output is based on the parameters that construct the lower triangular matrix from the Cholesky factorization of the covariance matrix of the random parameters. These parameters are labeled starting with _CHOL. For example, if two of the explanatory variables, $x1$ and $x2$, in your model have random coefficients, then the parameters that construct the diagonal of $\boldsymbol{\Gamma}$ are _CHOL.*x1.x1* and _CHOL.*x2.x2* and the lower part of $\boldsymbol{\Gamma}$ is _CHOL.*x1.x2*. If you use the NOCORR option, then the optimization is based on only the diagonal elements of $\boldsymbol{\Gamma}$, and in this case _CHOL.*x1.x1* and _CHOL.*x2.x2* are the standard deviations of the coefficients of $x1$ and $x2$, respectively. Although the optimization is performed with respect to $\boldsymbol{\theta}$, which includes the elements of $\boldsymbol{\Gamma}$ rather than $\boldsymbol{\Omega}$, the results are transformed to obtain the elements of $\boldsymbol{\Omega}$ and their corresponding standard errors.

Random-Effects Models

Random-effects models are a special case in which only the constant term is random. For these models, the parameter heterogeneity across individuals can be formulated as

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{0i} \\ \boldsymbol{\beta}_1 \end{pmatrix} = \begin{pmatrix} \beta_0 + \mu_i \\ \boldsymbol{\beta}_1 \end{pmatrix}$$

where μ_i has mean *0* and variance σ_μ^2.

In most applications of random-effects models, this type of parameter heterogeneity is modeled as a group-specific unobservable heterogeneity in the error term as

$$y_{it}^* = \mathbf{x}_{it}' \boldsymbol{\beta} + \epsilon_{it}$$

where

$$\epsilon_{it} = \mu_i + v_{it}$$

The density of an observed random variable, y_{it}, is

$$f(y_{it}|\mathbf{x}_{it}, \mu_i) = g(y_{it}, \mathbf{x}_{it}, \mu_i; \boldsymbol{\theta})$$

The density of the group-specific heterogeneity is

$$f(\mu_i) = h(\mu_i; \boldsymbol{\theta})$$

For example, in the case of a random-effects Tobit model, y_{it} is specified as

$$y^*_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \epsilon_{it}, \qquad t = 1, \ldots, T_i, \quad i = 1, \ldots, N$$

$$y_{it} = \begin{cases} y^*_{it} & \text{if} y^*_{it} > 0 \\ 0 & \text{if} y^*_{it} \leq 0 \end{cases}$$

where

$$\epsilon_{it} = \mu_i + v_{it}$$

$$v_{it}|(\mathbf{x}_i, \mu_i) \sim N(0, \sigma^2)$$

$$\mu_i|\mathbf{x}_i \sim N(0, \sigma^2_\mu)$$

where $\mathbf{x}_i$ contains $\mathbf{x}_{it}$ for all t and $\boldsymbol{\theta}$ consists of σ and σ_μ. Therefore, for this model,

$$f(y_{it}|\mathbf{x}_{it}, \mu_i) = \{1 - \Phi[(\mathbf{x}'_{it}\boldsymbol{\beta} + \mu_i)/\sigma]\}^{1[y_{it}=0]}\{(1/\sigma)\phi[(y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta} - \mu_i)/\sigma]\}^{1[y_{it}>0]}$$

and

$$f(\mu_i) = \phi(\mu_i/\sigma_\mu)$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution, $\phi(\boldsymbol{\omega}_i)$ is the probability density function of the standard normal distribution, and $1[\cdot]$ is the indicator function.

For random-effects models, the unobserved component, μ_i, must be integrated out in order to form the likelihood function for the observed data. For individual i,

$$L_i = f(y_{i1}, y_{i2}, \ldots, y_{iT_i}|\mathbf{x}_i, \boldsymbol{\beta}; \boldsymbol{\theta}) = \int_\mu \left[\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}'_{it}\boldsymbol{\beta}, \mu_i; \boldsymbol{\theta})\right] h(\mu_i; \boldsymbol{\theta}) d\mu_i$$

Therefore, the log-likelihood function for the observed data becomes

$$\ln L = \sum_{i=1}^{N} \ln \left[\int_\mu \left(\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}'_{it}\boldsymbol{\beta}, \mu_i; \boldsymbol{\theta})\right) h(\mu_i; \boldsymbol{\theta}) d\mu_i\right]$$

The notation for the likelihood function of a random-effects model is not much different from that of the random-parameters model discussed in the section "General Models with Random Parameters" on page 2009. However, there is a substantial difference in the formulation of the likelihood function of the random-parameters model. The integration in $\ln L$ is a multidimensional integral. More specifically, if the number of random parameters is K, then it is a K-dimensional integral.

Estimation

The integral in the log-likelihood function for random-parameters models does not have a closed form; that is, it is difficult to integrate out the random parameters. However, the integral can be approximated, and the usual likelihood estimation can be pursued based on the approximated log-likelihood function. PROC QLIM offers three methods of approximation: Monte Carlo (MC) integration, the quasi–Monte Carlo (QMC) method using the Halton sequences, and approximation by Hermite quadrature. The first two methods are simulation methods, and hence the likelihood method based on the resulting simulated log-likelihood function is called the simulated maximum likelihood. The third method fails to provide a good approximation when the dimensionality of the random parameters, K, is high. The Hermite quadrature method can be used only for random-effects models or random-parameters models that have a single random coefficient (that is, $K = 1$).

Monte Carlo Integration

Consider the random-effects model defined in the section "Random-Effects Models" on page 2011. First, note that

$$\int_{\mu} \left(\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}_{it}, \mu_i; \boldsymbol{\theta}) \right) h(\mu_i; \boldsymbol{\theta}) d\mu_i = E[F(\mu_i; \boldsymbol{\theta})]$$

The function is smooth, continuous, and continuously differentiable. By the law of large numbers, if $(\mu_{i1}, \mu_{i2}, \ldots, \mu_{iR})$ is a sample of iid draws from $h(\mu_i; \boldsymbol{\theta})$, then

$$\operatorname{plim} \frac{1}{R} \sum_{r=1}^{R} F(\mu_{ir}; \boldsymbol{\theta}) = E[F(\mu_i; \boldsymbol{\theta})]$$

This operation is implemented by simulation that uses a random number generator. PROC QLIM inserts the simulated integral in the log likelihood to obtain the simulated log likelihood

$$\ln L_{\text{simulated}} = \sum_{i=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \left(\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}_{it}, \mu_{ir}; \boldsymbol{\theta}) \right) \right]$$

and maximizes the simulated log likelihood with respect to the parameter set $\boldsymbol{\theta}$ that includes $\boldsymbol{\beta}$ and σ_μ.

Under certain assumptions (Greene 2001), the simulated likelihood estimator and the maximum likelihood estimator are equivalent. For this equivalence result to hold, the number of draws, R, must increase faster than the number of observations, N. For this reason, if the NDRAW= option is not specified, then by default, it is tied to the sample size by using the rule $R = N^{1+\delta}$, where $\delta = 1/2$.

Generalization of the log-likelihood function for random-parameters models is

$$\ln L_{\text{simulated}} = \sum_{i=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \left(\prod_{t=1}^{T_i} g(y_{it}, \boldsymbol{\beta}_{ir}, \mathbf{x}_{it}; \boldsymbol{\theta}) \right) \right]$$

where

$$\boldsymbol{\beta}_{ir} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 + \boldsymbol{\Gamma} \boldsymbol{\omega}_{ir} \end{pmatrix}$$

In this more general case, $\boldsymbol{\omega}_{ir}$ is the rth K-variate vector of random draws for individual i. The random draws come from the distribution with the probability density function $h(\boldsymbol{\omega}; \boldsymbol{\theta})$. PROC QLIM specifies $h(\boldsymbol{\omega}; \boldsymbol{\theta})$ as the probability density function of the standard normal distribution.

The use of independent random draws in simulation is conceptually straightforward, and the statistical properties of the simulated maximum likelihood estimator are easy to derive. However, simulation is a very computationally intensive technique. Moreover, the simulation method itself contributes to the variation of the simulated maximum likelihood estimator (see, for example, Geweke 1995). There are other ways to take draws that can provide greater accuracy by covering the domain of the integral more uniformly and by lowering the simulation variance (Train 2009, section 9.3). For example, quasi–Monte Carlo methods are based on an integration technique that replaces the pseudorandom draws of MC integration with a sequence of judiciously selected nonrandom points that provide more uniform coverage of the domain of the integral. Therefore, the advantage of QMC integration over MC integration is that for some types of sequences, the accuracy is far greater, convergence is much faster, and the simulation variance is smaller. QMC methods are surveyed in Bhat (2001), Sloan and Woźniakowski (1998), and Morokoff and Caflisch (1995). In addition to MC simulation, PROC QLIM offers the QMC integration method that uses Halton sequences.

QMC Method Using the Halton Sequence

Halton sequences (Halton 1960) provide uniform coverage for each observation's integral, and they decrease the simulation variance by inducing a negative correlation over the draws for each observation. A Halton sequence is constructed deterministically in terms of a prime number as its base. For example, the following sequence is the Halton sequence for 2:

$$1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, \ldots$$

For more information about how to generate a Halton sequence, see Train (2009), section 9.3.3.

If you use the QMC method, first, K Halton sequences are created—that is, one Halton sequence for each random parameter, with each sequence corresponding to a different prime number between 2 and the Kth prime number. Then for each sequence, part of the sequence (or the whole sequence, depending on whether you decide to discard the initial elements of the sequences[1]) is used in groups. For a given sequence, each group of consequent elements constitutes the "draws" for each cross-sectional observation. This way, each sub-sequence fills in the gaps for the previous sub-sequences, and the draws for one observation tend to be negatively correlated with those for the previous observation.

When the number of draws that are used for each observation rises, the coverage for each observation improves. This improvement in turn improves the accuracy; however, the negative covariance across observations diminishes. Because Halton draws are far more effective than random draws in Monte Carlo simulation, a small number of Halton draws provide relatively good integration (Spanier and Maize 1991).

The Halton draws are for a uniform density. PROC QLIM obtains ω_{ir} by evaluating the inverse cumulative standard normal density for each element of the rth K-variate draw for the ith group.

Approximation by Hermite Quadrature

Consider the random-effects model that is defined in the section "Random-Effects Models" on page 2011. This method is the Butler and Moffitt (1982) approach, which is based on models in which μ_i has a normal

[1] When sequences are created in multiple dimensions, the initial part of the series is usually eliminated because the initial terms of multiple Halton sequences are highly correlated. However, there is no such correlation for a single dimension.

distribution. If μ_i is normally distributed with zero mean, then

$$\int_{\mu} \left(\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}'_{it}\boldsymbol{\beta}, \mu_i; \boldsymbol{\theta}) \right) h(\mu_i; \boldsymbol{\theta}) d\mu_i$$
$$= \frac{1}{\sigma_\mu \sqrt{2\pi}} \int_{-\infty}^{+\infty} \prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}'_{it}\boldsymbol{\beta}, \mu_i; \boldsymbol{\theta}) \exp\left(\frac{-\mu_i^2}{2\sigma_\mu^2} \right) d\mu_i$$

Let $r_i = \mu_i/(\sigma_\mu\sqrt{2})$. Then $\mu_i = (\sigma_\mu\sqrt{2})r_i$ and $d\mu_i = (\sigma_\mu\sqrt{2})dr_i$. Making the change of variable and letting the error effects be additive produce

$$L_i = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp(-r_i^2) \left[\prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}'_{it}\boldsymbol{\beta} + (\sigma_\mu\sqrt{2})r_i; \boldsymbol{\theta}) \right] dr_i$$

This likelihood function is in a form that can be approximated accurately by using Gauss-Hermite quadrature, which eliminates the integration. Thus, the log-likelihood function can be approximated with

$$\ln L_h = \sum_{i=1}^{N} \ln \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} w_h \prod_{t=1}^{T_i} g(y_{it}, \mathbf{x}'_{it}\boldsymbol{\beta} + (\sigma_\mu\sqrt{2})r_h; \boldsymbol{\theta}) \right]$$

where w_h and r_h are the weights and nodes for the Hermite quadrature of degree H. PROC QLIM maximizes $\ln L_h$ when the Hermite quadrature option is specified (METHOD=HERMITE in the RANDOM statement).

Bayesian Analysis

To perform Bayesian analysis, you must specify a BAYES statement. Unless otherwise stated, all options in this section are options in the BAYES statement.

By default, PROC QLIM uses the random walk Metropolis algorithm to obtain posterior samples. For the implementation details of the Metropolis algorithm in PROC QLIM, such as the blocking of the parameters and tuning of the covariance matrices, see the sections "Blocking of Parameters" on page 2015 and "Tuning the Proposal Distribution" on page 2016.

The Bayes theorem states that

$$p(\theta|\mathbf{y}) \propto \pi(\theta)L(y|\theta)$$

where θ is a parameter or a vector of parameters and $\pi(\theta)$ is the product of the prior densities that are specified in the PRIOR statement. The term $L(y|\theta)$ is the likelihood associated with the MODEL statement.

Blocking of Parameters

In a multivariate parameter model, all the parameters are updated in one single block (by default or when you specify the SAMPLING=MULTIMETROPOLIS option). This could be inefficient, especially when parameters have vastly different scales. As an alternative, you could update the parameters one at the time (by specifying SAMPLING=UNIMETROPOLIS).

Tuning the Proposal Distribution

One key factor in achieving high efficiency of a Metropolis-based Markov chain is finding a good proposal distribution for each block of parameters. This process is called tuning. The tuning phase consists of a number of loops controlled by the options MINTUNE and MAXTUNE. The MINTUNE= option controls the minimum number of tuning loops and has a default value of 2. The MAXTUNE= option controls the maximum number of tuning loops and has a default value of 24. Each loop is iterated the number of times specified by the NTU= option, which has a default of 500. At the end of every loop, PROC QLIM examines the acceptance probability for each block. The acceptance probability is the percentage of NTU proposed values that have been accepted. If this probability does not fall within the acceptance tolerance range (see the following section), the proposal distribution is modified before the next tuning loop.

A good proposal distribution should resemble the actual posterior distribution of the parameters. Large sample theory states that the posterior distribution of the parameters approaches a multivariate normal distribution (see Gelman et al. 2004, Appendix B; Schervish 1995, Section 7.4). That is why a normal proposal distribution often works well in practice. The default proposal distribution in PROC QLIM is the normal distribution.

Scale Tuning

The acceptance rate is closely related to the sampling efficiency of a Metropolis chain. For a random walk Metropolis, a high acceptance rate means that most new samples occur right around the current data point. Their frequent acceptance means that the Markov chain is moving rather slowly and not exploring the parameter space fully. A low acceptance rate means that the proposed samples are often rejected; hence the chain is not moving much. An efficient Metropolis sampler has an acceptance rate that is neither too high nor too low. The scale c in the proposal distribution $q(\cdot|\cdot)$ effectively controls this acceptance probability. Roberts, Gelman, and Gilks (1997) show that if both the target and proposal densities are normal, the optimal acceptance probability for the Markov chain should be around 0.45 in a one-dimension problem and should asymptotically approach 0.234 in higher-dimension problems. The corresponding optimal scale is 2.38, which is the initial scale that is set for each block.

Because of the nature of stochastic simulations, it is impossible to fine-tune a set of variables so that the Metropolis chain has exactly the desired acceptance rate that you want. In addition, Roberts and Rosenthal (2001) empirically demonstrate that an acceptance rate between 0.15 and 0.5 is at least 80% efficient, so there is really no need to fine-tune the algorithms to reach an acceptance probability that is within a small tolerance of the optimal values. PROC QLIM works with a probability range, determined by TargetAcceptance ± 0.075. If the observed acceptance rate in a given tuning loop is less than the lower bound of the range, the scale is reduced; if the observed acceptance rate is greater than the upper bound of the range, the scale is increased. During the tuning phase, a scale parameter in the normal distribution is adjusted as a function of the observed acceptance rate and the target acceptance rate. PROC QLIM uses the following updating scheme,[2]

$$c_{\text{new}} = \frac{c_{\text{cur}} \cdot \Phi^{-1}(p_{\text{opt}}/2)}{\Phi^{-1}(p_{\text{cur}}/2)}$$

where c_{cur} is the current scale, p_{cur} is the current acceptance rate, and p_{opt} is the optimal acceptance probability.

[2] Roberts and associates demonstrate that the relationship between acceptance probability and scale in a random walk Metropolis scheme is $p = 2\Phi\left(-\sqrt{I}c/2\right)$, where c is the scale, p is the acceptance rate, Φ is the CDF of a standard normal, and $I \equiv E_f[(f'(x)/f(x))^2]$, $f(x)$ is the density function of samples (Roberts, Gelman, and Gilks 1997; Roberts and Rosenthal 2001). This relationship determines the updating scheme, with I replaced by the identity matrix to simplify calculation.

Covariance Tuning

To tune a covariance matrix, PROC QLIM takes a weighted average of the old proposal covariance matrix and the recent observed covariance matrix, based on the number samples (as specified by the NTU= option) NTU samples in the current loop. The formula to update the covariance matrix is

$$\mathrm{COV}_{\mathrm{new}} = 0.75\,\mathrm{COV}_{\mathrm{cur}} + 0.25\,\mathrm{COV}_{\mathrm{old}}$$

There are two ways to initialize the covariance matrix:

- The default is an identity matrix that is multiplied by the initial scale of 2.38 and divided by the square root of the number of estimated parameters in the model. A number of tuning phases might be required before the proposal distribution is tuned to its optimal stage, because the Markov chain needs to spend time to learn about the posterior covariance structure. If the posterior variances of your parameters vary by more than a few orders of magnitude, if the variances of your parameters are much different from 1, or if the posterior correlations are high, then the proposal tuning algorithm might have difficulty forming an acceptable proposal distribution.
- Alternatively, you can use a numerical optimization routine, such as the quasi-Newton method, to find a starting covariance matrix. The optimization is performed on the joint posterior distribution, and the covariance matrix is a quadratic approximation at the posterior mode. In some cases this is a better and more efficient way of initializing the covariance matrix. However, there are cases, such as when the number of parameters is large, where the optimization could fail to find a matrix that is positive definite. In those cases, the tuning covariance matrix is reset to the identity matrix.

A by-product of the optimization routine is that it also finds the maximum a posteriori (MAP) estimates with respect to the posterior distribution. The MAP estimates are used as the initial values of the Markov chain.

For more information, see the INIT statement.

Initial Values of the Markov Chains

You can assign initial values to any parameters. (For more information, see the INIT statement.) If you use the optimization option PROPCOV=, then PROC QLIM starts the tuning at the optimized values. This option overwrites the provided initial values. If you specify the RANDINIT option, the information that the INIT statement provides is overwritten.

Aggregation of Multiple Chains

When you want to exploit the possibility of running several MCMC instances at the same time (NTRDS=*n*>1), you face the problem of aggregating the chains. In ordinary applications, each MCMC instance can easily obtain stationary samples from the entire posterior distribution. In these applications, you can use the option AGGREGATION=UNWEIGHTED. This option piles up one chain on top of another and makes no particular adjustment. However, when the posterior distribution is characterized by multiple distinct posterior modes, some of the MCMC instances fail to obtain stationary samples from the entire posterior distribution. You can use the option AGGREGATION=WEIGHTED when the posterior samples from each MCMC instance approximate well only a part of the posterior distribution.

The main idea behind the option AGGREGATION=WEIGHTED is to consider the entire posterior distribution to be similar to a mixture distribution. When you are sampling with multiple threads, each MCMC instance

samples from one of the mixture components. Then the samples from each mixture component are aggregated together using a resampling scheme in which weights are proportional to the nonnormalized posterior distribution.

Description of the Algorithm

The preliminary step of the aggregation that is implied by the option AGGREGATION=WEIGHTED is to run several (K) independent instances of the MCMC algorithm. Each instance searches for a set of stationary samples. Notice that the concept of stationarity is weaker: each instance might be able to explore not the entire posterior but only portions of it. In the next equation, each column represents the output from one MCMC instance:

$$\begin{pmatrix} x_{11} \\ x_{21} \\ \cdots \\ x_{n1} \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \\ \cdots \\ x_{n2} \end{pmatrix} \cdots \begin{pmatrix} x_{1K} \\ x_{2K} \\ \cdots \\ x_{nK} \end{pmatrix} \sim \text{globally/locally sampled from the posterior}$$

If the length of each chain is less than n, you can augment the corresponding chain by subsampling the chain itself. Each chain is then sorted with respect to the nonnormalized posterior density: $\pi(x_{[1].}) \leq \pi(x_{[2].}) \leq \cdots \pi(x_{[n].})$. Therefore,

$$\begin{pmatrix} x_{11} \\ x_{21} \\ \cdots \\ x_{n1} \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \\ \cdots \\ x_{n2} \end{pmatrix} \cdots \begin{pmatrix} x_{1K} \\ x_{2K} \\ \cdots \\ x_{nK} \end{pmatrix} \rightarrow \begin{pmatrix} x_{[1]1} \\ x_{[2]1} \\ \cdots \\ x_{[n]1} \end{pmatrix} \begin{pmatrix} x_{[1]2} \\ x_{[2]2} \\ \cdots \\ x_{[n]2} \end{pmatrix} \cdots \begin{pmatrix} x_{[1]K} \\ x_{[2]K} \\ \cdots \\ x_{[n]K} \end{pmatrix}$$

The final step is to use a multinomial sampler to resample each row i with weights proportional to the nonnormalized posterior densities:

$$\widetilde{x}_{(i-1)K+1}, \widetilde{x}_{(i-1)K+2}, \ldots, \widetilde{x}_{(i-1)K+K} \sim \text{Multinom}\left[x_{[i]1}, x_{[i]2}, \ldots, x_{[i]K}; \pi(x_{[i]1}), \pi(x_{[i]2}), \ldots, \pi(x_{[i]K})\right]$$

The resulting posterior sample,

$$\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_K, \ldots, \widetilde{x}_{(i-1)K+1}, \widetilde{x}_{(i-1)K+2}, \ldots, \widetilde{x}_{(i-1)K+K}, \ldots, \widetilde{x}_{(n-1)K+1}, \widetilde{x}_{(n-1)K+2}, \ldots, \widetilde{x}_{nK}$$

is a good approximation of the posterior distribution that is characterized by multiple modes.

Automated Initialization of MCMC

The MCMC methods can generate samples from the posterior distribution. The correct implementation of these methods often requires the stationarity analysis, the convergence analysis and the accuracy analysis of the posterior samples. These analyses usually imply the following:

- initialization of the proposal distribution
- initialization of the chains (starting values)
- determination of the burn-in
- determination of the length of the chains.

In more general terms, this determination is equivalent to deciding whether the samples are drawn from the posterior distribution (stationarity analysis), and whether the number of samples is large enough to accurately approximate the posterior distribution (accuracy analysis). You can use the AUTOMCMC option to automate and facilitate the stationary analysis and the accuracy analysis.

Description of the Algorithm

The algorithm consists of two phases. In the first phase, the stationarity phase, the algorithm tries to generate stationary samples from the posterior distribution. In the second phase, the accuracy phase, the algorithm searches for an accurate representation of the posterior distribution. The algorithm implements the following tools:

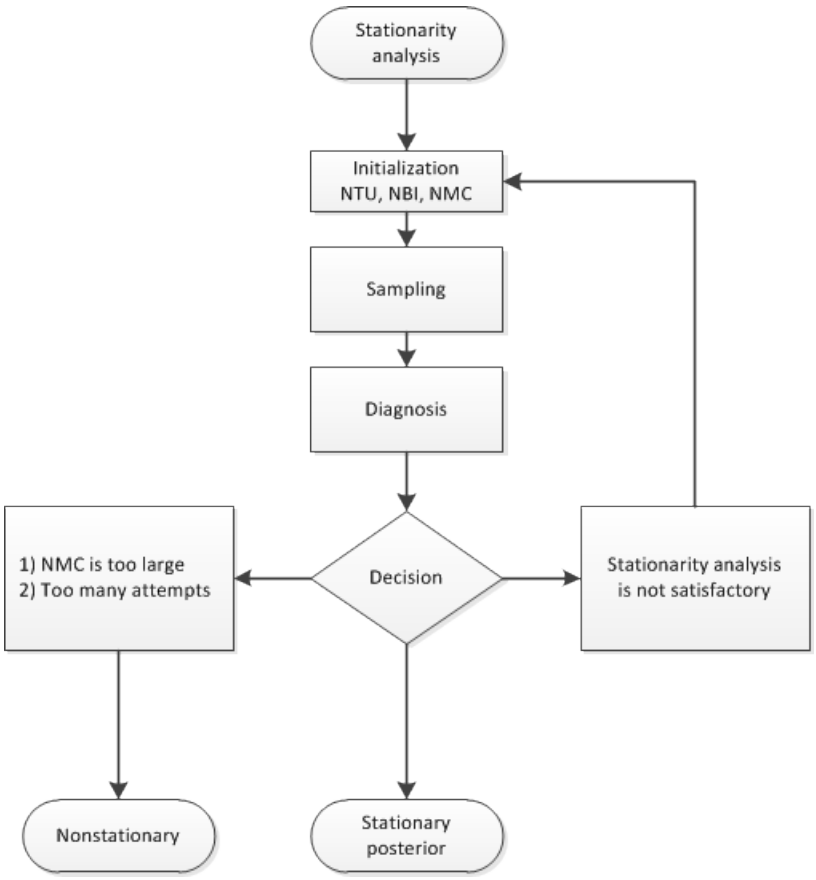
- Geweke test to check stationarity
- Heidelberger-Welch test to check stationarity and provide a proxy for the burn-in
- Heidelberger-Welch half-test to check the accuracy of the posterior mean
- Raftery-Lewis test to check the accuracy of a given percentile (indirectly proving a proxy for the number of required samples)
- effective sample size analysis to determine a proxy of the number of required samples

During the stationarity phase, the algorithm searches for stationarity. The number of attempts that the algorithm makes is determined by the option ATTEMPTS=*number*. During each attempt, a preliminary tuning stage chooses a proposal distribution for the MCMC sampler. At the end of the preliminary tuning phase, the algorithm analyzes tests for the stationarity of the samples. If the percentage of successful stationary tests is equal to or greater than the percentage that is indicated by the option TOL=*value*, then the posterior sample is considered to be stationary. If the sample cannot be considered stationary, then the algorithm attempts to achieve stationarity by changing some of the initialization parameters as follows:

- increasing the number of tuning samples (NTU)
- increasing the number of posterior samples (NMC)
- increasing the burn-in (NBI)

Figure 28.8 shows a flowchart of the algorithm as it searches for stationarity.

Figure 28.8 Flowchart of the AUTOMCMC Algorithm: Stationarity Analysis



You can initialize NMC=M, NBI=B, and NTU=T during the stationarity phase by specifying NMC, NBI, and NTU as options in the BAYES statement. You can also change the minimum stationarity acceptance ratio of successful stationarity tests that are needed to exit the stationarity phase. By default, TOL=0.95. For example:

```
proc qlim data=dataset;
   ...;
   bayes nmc=M nbi=B ntu=T automcmc=( stationarity=(tol=0.95) );
   ...;
run;
```

During the accuracy phase, the algorithm attempts to determine how many posterior samples are needed. The number of attempts is determined by the option ATTEMPTS=*number*. You can choose between two different approaches to study the accuracy:

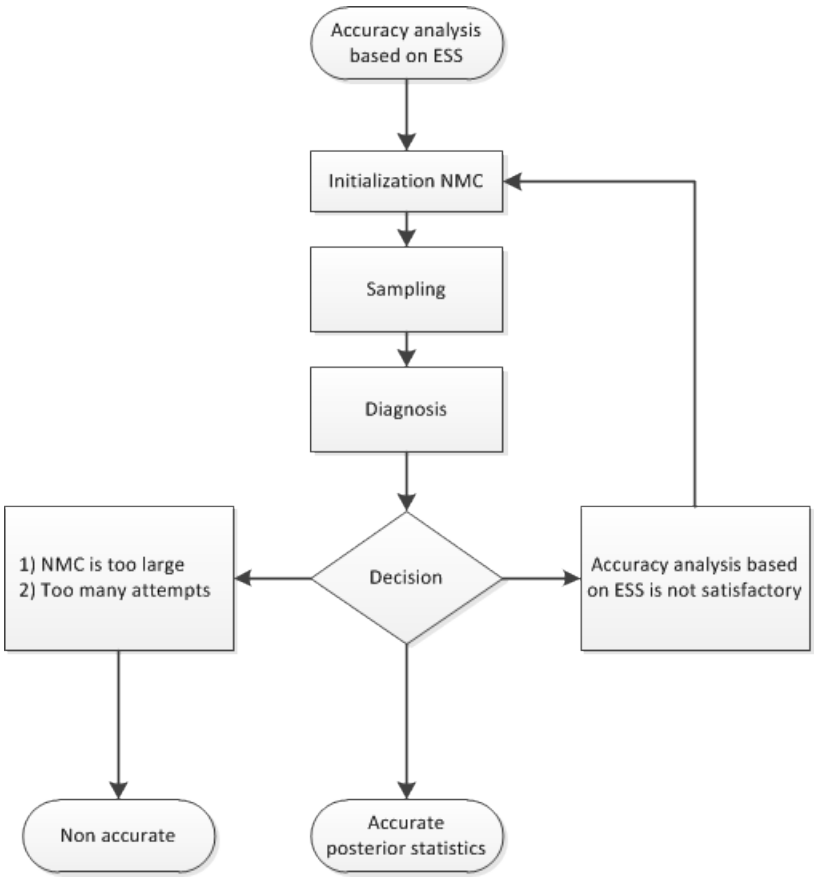
- accuracy analysis based on the effective sample size (ESS)
- accuracy analysis based on the Heidelberger-Welch half-test and the Raftery-Lewis test

If you choose the effective sample size approach, you must provide the minimum number of effective samples that are needed. You can also change the tolerance for the ESS accuracy analysis (by default, TOL=0.95). For example:

```
proc qlim data=dataset;
   ...;
   bayes automcmc=(targetess=N accuracy=(tol=0.95));
   ...;
run;
```

Figure 28.9 shows a flowchart of the algorithm based on the effective sample size approach to determine whether the samples provide an accurate representation of the posterior distribution.

Figure 28.9 Flowchart of the AUTOMCMC Algorithm: Accuracy Analysis Based on the ESS



If you choose the accuracy analysis based on the Heidelberger-Welch half-test and the Raftery-Lewis test (the default option), then you might want to choose a posterior quantile of interest for the Raftery-Lewis test (by default, 0.025). You can also change the tolerance for the accuracy analysis (by default, TOL=0.95). Notice that the Raftery-Lewis test produces a proxy of the number of posterior sample required. In each attempt, the current number of posterior samples is compared to this proxy. If the proxy is greater than the current nmc, then the algorithm reinitializes itself. To control this reinitialization, you can use the option RLLIMITS=(LB=lb UB=ub). In particular, there are three cases:

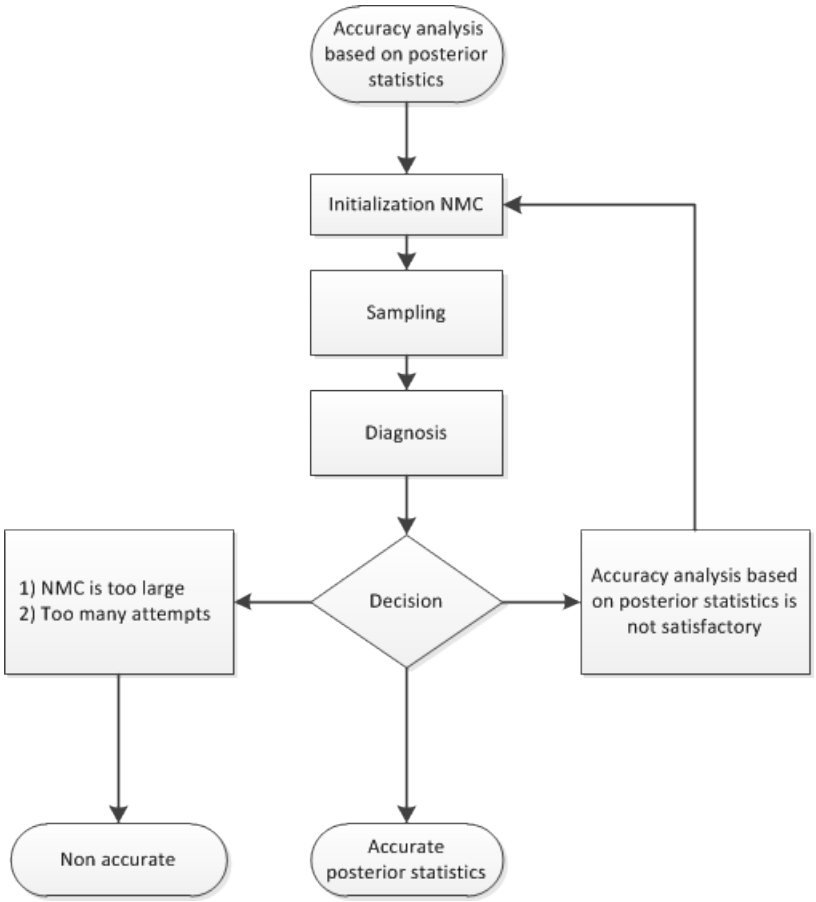
- If the proxy is greater than ub, then NMC is set equal to ub.
- If the proxy is less than lb, then NMC is set equal to lb.
- If lb is less than the proxy, which is less than ub, then NMC is set equal to the proxy.

For example:

```
proc qlim data=dataset;
   ...;
   bayes automcmc=( accuracy=(tol=0.95 targetstats=(rllimits=(lb=k1 ub=k2))) )
        raftery(q=0.025);
   ...;
run;
```

Figure 28.10 shows a flowchart of the algorithm based on the Heidelberger-Welch half-test and the Raftery-Lewis test approach to determine whether the posterior samples provide an accurate representation of the posterior distribution.

Figure 28.10 Flowchart of the AUTOMCMC Algorithm: Accuracy Analysis Based on the Heidelberger-Welch Half-Test and the Raftery-Lewis Test



Prior Distributions

The PRIOR statement is used to specify the prior distribution of the model parameters. You must specify a list of parameters, a tilde ∼, and then a distribution with its parameters. You can specify multiple PRIOR statements to define independent priors. Parameters that are associated with a regressor variable are referred to by the name of the corresponding regressor variable.

You can specify the special keyword _REGRESSORS to consider all the regressors of a model. If multiple prior statements affect the same parameter, the prior that is specified is used. For example, in a regression with three regressors (X1, X2, X3) the following statements imply that the prior on X1 is NORMAL(MEAN=0, VAR=1), the prior on X2 is GAMMA(SHAPE=3, SCALE=4), and the prior on X3 is UNIFORM(MIN=0, MAX=1):

```
...
prior _Regressors ~ uniform(min=0, max=1);
prior X1 X2 ~ gamma(shape=3, scale=4);
prior X1 ~ normal(mean=0, var=1);
...
```

If a parameter is not associated with a PRIOR statement or if some of the prior hyperparameters are missing, then the default choices shown in Table 28.2 are considered.

Table 28.2 Default Values for Prior Distributions

PRIOR *distribution*	Hyperparameter$_1$	Hyperparameter$_2$	Min	Max	Parameters Default Choice
NORMAL	MEAN=0	VAR=1E6	$-\infty$	∞	Regression-Location-Threshold
IGAMMA	SHAPE=2.000001	SCALE=1	> 0	∞	Scale
SQIGAMMA	SHAPE=2.000001	SCALE=1	> 0	∞	Scale
GAMMA	SHAPE=1	SCALE=1	0	∞	
SQGAMMA	SHAPE=1	SCALE=1	0	∞	
UNIFORM			$-\infty$	∞	
UNIFORM			> -1	< 1	Cross-correlation
BETA	SHAPE1=1	SHAPE2=1	$-\infty$	∞	
T	LOCATION=0	DF=3	$-\infty$	∞	

For density specification, see the section "Standard Distributions" on page 2029.

Priors for Heteroscedastic Models

The choice of the prior distribution for a heteroscedastic model is particularly interesting. Based on the notation provided in section "HETERO Statement" on page 1975, you need to provide a prior for $\boldsymbol{\gamma}$. This prior is enough to induce different σ_i^2 into the analysis. The resulting inference is a compromise between two cases: the inference based on the entire sample and the inference based on a single unit $\mathbf{z}_i$. The degree of compromise is determined by $\pi(\boldsymbol{\gamma})$.

This type of modeling is similar to a method called "hierarchical Bayes," in which the prior is characterized by two levels: one for each individual $\pi(\sigma_i^2|\boldsymbol{\gamma})$ and one for the entire population $\pi(\boldsymbol{\gamma})$. In this scenario the degree of compromise between the information provided by a unit and the information provided by the entire sample is determined by the data.

The choice of the prior might not be straightforward, and it can heavily affect sampling performance.

Depending on how the heteroscedastic effects are modeled, the default priors are

$$\text{if } \left[1+\exp(\mathbf{z}_i'\boldsymbol{\gamma})\right], \quad \pi(\gamma_j) = \textbf{normal}\left\{\text{mean} = \frac{1}{\bar{z}_j J}\left[\log\left(\frac{\varepsilon^4}{1+\varepsilon^2}\right)\right], \text{var} = \frac{1}{\bar{z}_j^2 J}\left[\log\left(\frac{1+\varepsilon^2}{\varepsilon^2}\right)\right]\right\}$$

$$\text{if } \left[\exp(\mathbf{z}_i'\boldsymbol{\gamma})\right], \quad \pi(\gamma_j) = \textbf{normal}\left\{\text{mean} = \frac{1}{\bar{z}_j J}\left[\log\left(\frac{1}{2}\right)\right], \text{var} = \frac{1}{\bar{z}_j^2 J}\left[\log(2)\right]\right\}$$

$$\text{if } \left(1+\mathbf{z}_i'\boldsymbol{\gamma}\right), \quad \pi(\gamma_j) = \textbf{normal}\left\{\text{mean} = 0, \text{var} = \frac{1}{\bar{z}_j^2 J}\right\}$$

$$\text{if } \left[1+(\mathbf{z}_i'\boldsymbol{\gamma})^2\right], \quad \pi(\gamma_j) = \textbf{normal}\left\{\text{mean} = \frac{(\varepsilon^2-1/2)^{1/4}}{\bar{z}_j J}, \text{var} = \frac{\varepsilon-(\varepsilon^2-1/2)^{1/2}}{\bar{z}_j^2 J}\right\}$$

where $\bar{z}_j = \frac{1}{n}\sum_{i=1}^{n} z_{ij}$, $\forall j$, and ε is a small number (by default, $\varepsilon = 0.1$ for the EXPONENTIAL link function and $\varepsilon = 0.71$ for the QUADRATIC link function).

The priors for the EXPONENTIAL and QUADRATIC link functions are not straightforward. To understand the choices, do the following:

1. Assume that
$$\mathbf{z}_i'\boldsymbol{\gamma} = z_{i1}\gamma_1 + \cdots + z_{iJ}\gamma_J \approx \bar{z}_1\gamma_1 + \cdots + \bar{z}_J\gamma_J, \qquad \forall i$$

2. Set the priors according to the link function type:

   - For the EXPONENTIAL link function, set
   $$\begin{aligned} \mathrm{E}\left[\exp(\mathbf{z}_i'\boldsymbol{\gamma})\right] &\approx \mathrm{E}\left[\exp(\bar{z}_1\gamma_1)\right] \times \cdots \times \mathrm{E}\left[\exp(\bar{z}_J\gamma_J)\right] = \varepsilon \\ \mathrm{V}\left[\exp(\mathbf{z}_i'\boldsymbol{\gamma})\right] &\approx \mathrm{E}\left[\exp(2\bar{z}_1\gamma_1)\right] \times \cdots \times \mathrm{E}\left[\exp(2\bar{z}_J\gamma_J)\right] - \varepsilon^2 = 1 \end{aligned}$$
   Assume a normal prior for $\pi(\gamma_j)$, and set
   $$\begin{aligned} \mathrm{E}\left[\exp(\bar{z}_j\gamma_j)\right] &= \varepsilon^{\frac{1}{J}}, \forall j \\ \mathrm{E}\left[\exp(2\bar{z}_j\gamma_j)\right] &= (1+\varepsilon^2)^{\frac{1}{J}}, \forall j \end{aligned}$$
   Based on the properties of the lognormal distribution, the prior hyperparameters for γ_j can be derived. Notice that J is the number of regressors that are used in the heterogeneous regression. If the intercept is excluded, then $\varepsilon = 1$.

   - For the QUADRATIC link function, set
   $$\begin{aligned} \mathrm{E}\left[(\mathbf{z}_i'\boldsymbol{\gamma})^2\right] &\approx \left[\mathrm{E}\left(\bar{z}_1\gamma_1 + \cdots + \bar{z}_J\gamma_J\right)\right]^2 + \mathrm{V}\left[\bar{z}_1\gamma_1 + \cdots + \bar{z}_J\gamma_J\right] = \varepsilon \\ \mathrm{V}\left[(\mathbf{z}_i'\boldsymbol{\gamma})^2\right] &\approx \mathrm{E}\left[(\bar{z}_1\gamma_1 + \cdots + \bar{z}_J\gamma_J)^4\right] - \varepsilon^2 = 1 \end{aligned}$$
   Assume a normal prior for $\pi(\gamma_j)$. Based on the properties of the normal distribution, the preceding expressions return
   $$\begin{aligned} \mathrm{E}\left[\bar{z}_1\gamma_1 + \cdots + \bar{z}_J\gamma_J\right] &= (\varepsilon^2-1/2)^{1/4} \\ \mathrm{V}\left[\bar{z}_1\gamma_1 + \cdots + \bar{z}_J\gamma_J\right] &= \varepsilon - (\varepsilon^2-1/2)^{1/2} \\ \varepsilon &> (1/2)^{1/2} \end{aligned}$$

The prior hyperparameters for γ_j can be derived by setting

$$\begin{aligned} \mathrm{E}\left[\bar{z}_j \gamma_j\right] &= \frac{(\varepsilon^2 - 1/2)^{1/4}}{J}, \forall j \\ \mathrm{V}\left[\bar{z}_j \gamma_j\right] &= \frac{\varepsilon - (\varepsilon^2 - 1/2)^{1/2}}{J}, \forall j \end{aligned}$$

Notice that J is the number of regressors that are used in the heterogeneous regression. It is important to emphasize that the restriction $\varepsilon > (1/2)^{1/2} \approx 0.71$ is likely to introduce some distortion because ε cannot be any "small" number.

Automated MCMC

The main purpose is to provide the user with the opportunity of obtaining a good approximation of the posterior distribution without initializing the MCMC algorithm: initial values, proposal distributions, burn-in and number of samples.

The automated algorithm is composed of two phases: tuning and sampling. In the tuning phase, there are two main concerns: the choice of a good proposal distribution and the search for the stationary region of the posterior distribution. In the sampling phase, the algorithm will decide how many samples are necessary to obtain good approximations of the posterior mean and some quantiles of interest.

Stationarity Phase

During the stationarity phase, the algorithm tries to search for a good proposal distribution and, at the same time, to reach the stationary region of the posterior. The choice of the proposal distribution is based on the analysis of the acceptance rates. This is similar to what is done in PROC MCMC; for more information, see the section "Tuning the Proposal Distribution" (Chapter 74, *SAS/STAT User's Guide*). For the stationarity analysis, the main idea is to run two tests, Geweke (Ge) and Heidelberger-Welch (HW), on the posterior chains at the end of each attempt. For more information, see the sections "Geweke Diagnostics" (Chapter 7, *SAS/STAT User's Guide*) and "Heidelberger and Welch Diagnostics" (Chapter 7, *SAS/STAT User's Guide*). If the stationarity hypothesis is rejected, then the tuning samples are increased and the tests repeated in the next attempt. After 10 attempts, the stationarity phase will be ended regardless of the results. The tuning parameters for the first attempt are fixed:

1000	burn-in (nbi)
500	tuning samples (ntu)
1000	MCMC samples (nmc)

For the remaining attempts, the tuning parameters will be adjusted dynamically. More specifically, each parameter will be assigned an acceptance ratio (AR) of the stationarity hypothesis,

$AR_i = 0$	if	both tests reject the stationarity hypothesis
$AR_i = 0.5$	if	one tests rejects and the other does not
$AR_i = 1$	if	both tests do not reject the stationarity hypothesis

for $i = 1, \ldots, k$. For the Geweke test, the implemented significance level is 0.05. Then, an overall stationarity average (SA) for all parameters ratios is evaluated,

$$SA = \sum_{i=1}^{k} \frac{AR_i}{k}$$

and the number of tuning samples is updated accordingly:

$$\begin{aligned}
&\text{ntu} = \text{ntu} + 2000 \quad &&\text{if} \quad &&SA < 70\% \\
&\text{ntu} = \text{ntu} + 1000 \quad &&\text{if} \quad &&70\% \leq SA < 100\% \\
&\text{ntu} = \text{ntu} \quad &&\text{if} \quad &&SA = 100\%
\end{aligned}$$

The burn-in is also updated whenever stationarity is not achieved:

$$\text{nbi} = \text{nbi} + 1000$$

Moreover, the Heidelberger-Welch test also provides an indications of how much burn-in should be used. The algorithm requires this burn-in to be nbi(HW) $= 0$. If that is not the case, the burn-in will updated accordingly,

$$\text{nbi} = \max[\text{nbi}, \text{nbi(HW)}]$$

and a new attempt searching for stationarity will be implemented. This choice is motivated by the fact that the burn-in must be discarded in order to reach the stationary region of the posterior distribution.

The number of samples is updated at each attempt. However, in order to exit the stationarity phase, it will not be required nmc(RL) $= 0$. The default update is nmc $=$ nmc $+ 1000$. Depending on the outcome of the Raftery-Lewis diagnostics, if nmc $< \min \{\text{LB}[\text{nmc(RL)}], \text{nmc(RL)}\}$, the number of sampling is further updated to nmc $= \text{LB}[\text{nmc(RL)}]$. By default, LB [nmc(RL)] $= 10000$. Finally, if the number of projected samples is not sufficient to perform a stable evaluation of the Raftery-Lewis test, the number of samples is updated to nmc $= \min[\text{nmc(RL)}]$. For more information, see the section "AUTOMCMC<=(*automcmc-options*)>" on page 1964 and Chapter 7.4, "Raftery and Lewis Diagnostics" (*SAS/STAT User's Guide*).

Accuracy Phase

The main idea of the accuracy phase is to make sure that the mean and a quantile of interest are evaluated accurately. This can be tested by implementing the half-width test by Heidelberger-Welch and by analyzing the Raftery-Lewis diagnostic tool. In addition, the requirements defined in the stationarity phase will also be checked: the Geweke and the Heidelberger-Welch tests must not reject the stationary hypothesis and the burn-in predicted by the Heidelberger-Welch test must be zero.

The accuracy phase is characterized by a maximum of 10 attempts. If the algorithm exceeds this limit, the accuracy phase will end and indications on how to improve sampling will be given. The search of accuracy can be performed using two different method. The first method (the default) is triggered by the option TARGETSTATS and it is based on the accuracy analysis of the mean and a percentile of interest. The second method is triggered by the option TARGETESS and it targets a minimum number of effective samples. The accuracy phase will first update the burn-in with the information provided by the HW test: nbi $=$ nbi $+$ nbi(HW). Then, it determines the difference between the actual number of samples

and the number of samples predicted by either the RL test or the ESS: $\Delta[\text{nmc}] = \text{nmc(RL)} - \text{nmc}$, or $\Delta[\text{nmc}] = \text{nmc(ESS)} - \text{nmc}$. The new number of samples will be updated accordingly:

$$\begin{aligned}
&\text{nmc} = \text{nmc} + \text{LB}\,[\text{nmc(RL)}] && \text{if} && 0 < \Delta[\text{nmc}] \leq \text{LB}\,[\text{nmc(RL)}] \\
&\text{nmc} = \text{nmc} + \Delta[\text{nmc}] && \text{if} && \text{LB}\,[\text{nmc(RL)}] < \Delta[\text{nmc}] \leq \text{UB}\,[\text{nmc(RL)}] \\
&\text{nmc} = \text{nmc} + \text{UB}\,[\text{nmc(RL)}] && \text{if} && \text{UB}\,[\text{nmc(RL)}] < \Delta[\text{nmc}]
\end{aligned}$$

By default, $\text{LB}\,[\text{nmc(RL)}] = 10000$ and $\text{UB}\,[\text{nmc(RL)}] = 300000$.

In addition, the accuracy search triggered by the option TARGETSTATS also implements the HW half-width test to checks whether the sample mean is accurate. If the mean of any parameters is not considered to be accurate and the number of samples has not been updated based on $\Delta[\text{nmc}]$, then the number of samples is increased:

$$\text{nmc} = \text{nmc} + 5000 \quad \text{if} \quad \Delta[\text{nmc}] \leq 0$$

Marginal Likelihood

The Bayes theorem states that

$$p(\theta|\mathbf{y}) \propto \pi(\theta)L(y|\theta)$$

where θ is a vector of parameters and $\pi(\theta)$ is the product of the prior densities that are specified in the PRIOR statement. The term $L(y|\theta)$ is the likelihood that is associated with the MODEL statement. The function $\pi(\theta)L(y|\theta)$ is the nonnormalized posterior distribution over the parameter vector θ. The normalized posterior distribution (simply, the posterior distribution) is

$$p(\theta|\mathbf{y}) = \frac{\pi(\theta)L(y|\theta)}{\int_\theta \pi(\theta)L(y|\theta)d\theta}$$

The denominator $m(y) = \int_\theta \pi(\theta)L(y|\theta)d\theta$ (also called the "marginal likelihood") is a quantity of interest because it represents the probability of the data after the effect of the parameter vector has been averaged out. Because of its interpretation, the marginal likelihood can be used in various applications, including model averaging, variable selection, and model selection.

A natural estimate of the marginal likelihood is provided by the harmonic mean,

$$m(y) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{L(y|\theta_i)} \right\}^{-1}$$

where θ_i is a sample draw from the posterior distribution. In practical applications, this estimator has proven to be unstable.

An alternative and more stable estimator can be obtained with an importance sampling scheme. The auxiliary distribution for the importance sampler can be chosen through the cross entropy theory (Chan and Eisenstat 2015). In particular, given a parametric family of distributions, the auxiliary density function is chosen to be the one closest, in terms of the Kullback-Leibler divergence, to the probability density that would give a zero variance estimate of the marginal likelihood. In practical terms, this is equivalent to the following algorithm:

1. Choose a parametric family, $f(., \beta)$, for the parameters of the model: $f(\theta|\beta)$.
2. Evaluate the maximum likelihood estimator of β by using the posterior samples $\theta_1, \ldots, \theta_n$ as data.
3. Use $f(\theta^*|\hat{\beta}_{mle})$ to generate the importance samples $\theta_1^*, \ldots, \theta_{n^*}^*$.
4. Estimate the marginal likelihood:

$$m(y) = \frac{1}{n^*} \sum_{j=1}^{n^*} \frac{L(y|\theta_j^*)\pi(\theta_j^*)}{f(\theta_j^*|\hat{\beta}_{mle})}$$

The parametric family for the auxiliary distribution is chosen to be Gaussian. The parameters that are subject to bounds are transformed accordingly:

- If $-\infty < \theta < \infty$, then $p = \theta$.
- If $m \leq \theta < \infty$, then $q = \log(\theta - m)$.
- If $-\infty < \theta \leq M$, then $r = \log(M - \theta)$.
- If $m \leq \theta \leq M$, then $s = \log(\theta - m) - \log(M - \theta)$.

Assuming independence for the parameters that are subject to bounds, the auxiliary distribution to generate importance samples is

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \\ \mathbf{r} \\ \mathbf{s} \end{pmatrix} \sim \mathbf{N}\left[\begin{pmatrix} \mu_p \\ \mu_q \\ \mu_r \\ \mu_s \end{pmatrix}, \begin{pmatrix} \Sigma_p & 0 & 0 & 0 \\ 0 & \Sigma_q & 0 & 0 \\ 0 & 0 & \Sigma_r & 0 \\ 0 & 0 & 0 & \Sigma_r \end{pmatrix} \right]$$

where $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$, and $\mathbf{s}$ are vectors that contain the transformations of the unbounded, bounded-below, bounded-above, and bounded-above-and-below parameters. Also, given the imposed independence structure, Σ_p can be a nondiagonal matrix, but Σ_q, Σ_r, and Σ_s are assumed to be diagonal matrices.

Standard Distributions

Table 28.3 through Table 28.10 show all the distribution density functions that PROC QLIM recognizes. You specify these distribution densities in the PRIOR statement.

Table 28.3 Beta Distribution

PRIOR statement	BETA(SHAPE1=a, SHAPE2=b, MIN=m, MAX=M) Note: Commonly $m = 0$ and $M = 1$.
Density	$\frac{(\theta-m)^{a-1}(M-\theta)^{b-1}}{B(a,b)(M-m)^{a+b-1}}$
Parameter restriction	$a > 0, \quad b > 0, \quad -\infty < m < M < \infty$
Range	$\begin{cases} [m, M] & \text{when } a = 1, b = 1 \\ [m, M) & \text{when } a = 1, b \neq 1 \\ (m, M] & \text{when } a \neq 1, b = 1 \\ (m, M) & \text{otherwise} \end{cases}$
Mean	$\frac{a}{a+b} \times (M - m) + m$
Variance	$\frac{ab}{(a+b)^2(a+b+1)} \times (M - m)^2$
Mode	$\begin{cases} \frac{a-1}{a+b-2} \times M + \frac{b-1}{a+b-2} \times m & a > 1, b > 1 \\ m \text{ and } M & a < 1, b < 1 \\ m & \begin{cases} a < 1, b \geq 1 \\ a = 1, b > 1 \end{cases} \\ M & \begin{cases} a \geq 1, b < 1 \\ a > 1, b = 1 \end{cases} \\ \text{not unique} & a = b = 1 \end{cases}$
Defaults	SHAPE1=SHAPE2=1, MIN $\rightarrow -\infty$, MAX $\rightarrow \infty$

Table 28.4 Gamma Distribution

PRIOR statement	GAMMA(SHAPE=a, SCALE=b)
Density	$\frac{1}{b^a \Gamma(a)} \theta^{a-1} e^{-\theta/b}$
Parameter restriction	$a > 0, b > 0$
Range	$[0, \infty)$
Mean	ab
Variance	ab^2
Mode	$(a - 1)b$
Defaults	SHAPE=SCALE=1

Table 28.5 Inverse-Gamma Distribution

PRIOR statement	IGAMMA(SHAPE=a, SCALE=b)
Density	$\frac{b^a}{\Gamma(a)}\theta^{-(a+1)}e^{-b/\theta}$
Parameter restriction	$a > 0, b > 0$
Range	$0 < \theta < \infty$
Mean	$\frac{b}{a-1}, \quad a > 1$
Variance	$\frac{b^2}{(a-1)^2(a-2)}, \quad a > 2$
Mode	$\frac{b}{a+1}$
Defaults	SHAPE=2.000001, SCALE=1

Table 28.6 Normal Distribution

PRIOR statement	NORMAL(MEAN=μ, VAR=σ^2)
Density	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right)$
Parameter restriction	$\sigma^2 > 0$
Range	$-\infty < \theta < \infty$
Mean	μ
Variance	σ^2
Mode	μ
Defaults	MEAN=0, VAR=1000000

Table 28.7 Square Root Gamma Distribution

PRIOR statement	SQGAMMA(SHAPE=a, SCALE=b)
Density	$\frac{2}{b^a\Gamma(a)}\theta^{2a-1}e^{-\theta^2/b}$
Parameter restriction	$a > 0, b > 0$
Range	$[0, \infty)$
Mean	$\frac{\Gamma(a+\frac{1}{2})}{\Gamma(a)}\sqrt{b}$
Variance	$\left\{a - \left[\frac{\Gamma(a+\frac{1}{2})}{\Gamma(a)}\right]^2\right\} b$
Mode	$\sqrt{(a-\frac{1}{2})b}, \quad a \geq \frac{1}{2}$
Defaults	SHAPE=SCALE=1

For more information, see Stacy (1962).

Table 28.8 Square Root Inverse-Gamma Distribution

PRIOR statement	SQIGAMMA(SHAPE=a, SCALE=b)
Density	$\frac{2b^a}{\Gamma(a)}\theta^{-(2a+1)}e^{-b/\theta^2}$
Parameter restriction	$a > 0, b > 0$
Range	$0 < \theta < \infty$
Mean	$\frac{\Gamma(a-\frac{1}{2})}{\Gamma(a)}\sqrt{b}, \qquad a > \frac{1}{2}$
Variance	$\left\{\frac{1}{a-1} - \left[\frac{\Gamma(a-\frac{1}{2})}{\Gamma(a)}\right]^2\right\} b, \qquad a > 1$
Mode	$\sqrt{\frac{b}{a+\frac{1}{2}}}$
Defaults	SHAPE=2.000001, SCALE=1

For more information, see Stacy (1962).

Table 28.9 *t* Distribution

PRIOR statement	T(LOCATION=μ, DF=ν)
Density	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left[1 + \frac{(\theta-\mu)^2}{\nu}\right]^{-\frac{\nu+1}{2}}$
Parameter restriction	$\nu > 0$
Range	$-\infty < \theta < \infty$
Mean	μ, for $\nu > 1$
Variance	$\frac{\nu}{\nu-2}$, for $\nu > 2$
Mode	μ
Defaults	LOCATION=0, DF=3

Table 28.10 Uniform Distribution

PRIOR statement	UNIFORM(MIN=m, MAX=M)
Density	$\frac{1}{M-m}$
Parameter restriction	$-\infty < m < M < \infty$
Range	$\theta \in [m, M]$
Mean	$\frac{m+M}{2}$
Variance	$\frac{(M-m)^2}{12}$
Mode	Not unique
Defaults	MIN$\rightarrow -\infty$, MAX$\rightarrow \infty$

Output to SAS Data Set

XBeta, Predicted, Residual

Xbeta is the structural part on the right-hand side of the model. Predicted value is the predicted dependent variable value. For censored variables, if the predicted value is outside the boundaries, it is reported as the closest boundary. For discrete variables, it is the level whose boundaries Xbeta falls between. Residual is defined only for continuous variables and is defined as

$$Residual = Observed - Predicted$$

Error Standard Deviation

Error standard deviation is σ_i in the model. It varies only when the HETERO statement is used.

Marginal Effects

Marginal effect is defined as a contribution of one control variable to the response variable. For the binary choice model with two response categories, $\mu_0 = -\infty$, $\mu_1 = 0$, $\mu_2 = \infty$; and ordinal response model with M response categories, $\mu_0, \ldots, \mu_M$, define

$$R_{i,j} = \mu_j - \mathbf{x}_i'\boldsymbol{\beta}$$

The probability that the unobserved dependent variable is contained in the *j*th category can be written as

$$P[\mu_{j-1} < y_i^* \le \mu_j] = F(R_{i,j}) - F(R_{i,j-1})$$

The marginal effect of changes in the regressors on the probability of $y_i = j$ is then

$$\frac{\partial \mathrm{Prob}[y_i = j]}{\partial \mathbf{x}} = [f(\mu_{j-1} - \mathbf{x}_i'\boldsymbol{\beta}) - f(\mu_j - \mathbf{x}_i'\boldsymbol{\beta})]\boldsymbol{\beta}$$

where $f(x) = \frac{dF(x)}{dx}$. In particular,

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} & (\text{probit}) \\ \frac{e^{-x}}{[1+e^{(-x)}]^2} & (\text{logit}) \end{cases}$$

The marginal effects in the Box-Cox regression model are

$$\frac{\partial E[y_i]}{\partial \mathbf{x}} = \boldsymbol{\beta} \frac{x^{\lambda_k - 1}}{y^{\lambda_0 - 1}}$$

The marginal effects in the truncated regression model are

$$\frac{\partial E[y_i | L_i < y_i^* < R_i]}{\partial \mathbf{x}} = \boldsymbol{\beta} \left[1 - \frac{(\phi(a_i) - \phi(b_i))^2}{(\Phi(b_i) - \Phi(a_i))^2} + \frac{a_i \phi(a_i) - b_i \phi(b_i)}{\Phi(b_i) - \Phi(a_i)} \right]$$

where $a_i = \frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i}$ and $b_i = \frac{R_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i}$.

The marginal effects in the censored regression model are

$$\frac{\partial E[y | \mathbf{x}_i]}{\partial \mathbf{x}} = \boldsymbol{\beta} \times \mathrm{Prob}[L_i < y_i^* < R_i]$$

Inverse Mills Ratio, Expected and Conditionally Expected Values

Expected and conditionally expected values are computed only for continuous variables. The inverse Mills ratio is computed for censored or truncated continuous, binary discrete, and selection endogenous variables.

Let L_i and R_i be the lower boundary and upper boundary, respectively, for the y_i. Define $a_i = \frac{L_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i}$ and $b_i = \frac{R_i - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma_i}$. Then the inverse Mills ratio is defined as

$$\lambda = \frac{(\phi(a_i) - \phi(b_i))}{(\Phi(b_i) - \Phi(a_i))}$$

for a continuous variable and defined as

$$\lambda = \frac{\phi(\mathbf{x}_i'\boldsymbol{\beta})}{\Phi(\mathbf{x}_i'\boldsymbol{\beta})}$$

for a binary discrete variable.

The expected value is the unconditional expectation of the dependent variable. For a censored variable, it is

$$E[y_i] = \Phi(a_i)L_i + (\mathbf{x}_i'\boldsymbol{\beta} + \lambda\sigma_i)(\Phi(b_i) - \Phi(a_i)) + (1 - \Phi(b_i))R_i$$

For a left-censored variable ($R_i = \infty$), this formula is

$$E[y_i] = \Phi(a_i)L_i + (\mathbf{x}_i'\boldsymbol{\beta} + \lambda\sigma_i)(1 - \Phi(a_i))$$

where $\lambda = \frac{\phi(a_i)}{1-\Phi(a_i)}$.

For a right-censored variable ($L_i = -\infty$), this formula is

$$E[y_i] = (\mathbf{x}_i'\boldsymbol{\beta} + \lambda\sigma_i)\Phi(b_i) + (1 - \Phi(b_i))R_i$$

where $\lambda = -\frac{\phi(b_i)}{\Phi(b_i)}$.

For a noncensored variable, this formula is

$$E[y_i] = \mathbf{x}_i'\boldsymbol{\beta}$$

The conditional expected value is the expectation given that the variable is inside the boundaries:

$$E[y_i | L_i < y_i < R_i] = \mathbf{x}_i'\boldsymbol{\beta} + \lambda\sigma_i$$

Probability

Probability applies only to discrete responses. It is the marginal probability that the discrete response is taking the value of the observation. If the PROBALL option is specified, then the probability for all of the possible responses of the discrete variables is computed.

Technical Efficiency

Technical efficiency for each producer is computed only for stochastic frontier models.

In general, the stochastic production frontier can be written as

$$y_i = f(x_i;\beta)\exp\{v_i\}TE_i$$

where y_i denotes producer i's actual output, $f(\cdot)$ is the deterministic part of production frontier, $\exp\{v_i\}$ is a producer-specific error term, and TE_i is the technical efficiency coefficient, which can be written as

$$TE_i = \frac{y_i}{f(x_i;\beta)\exp\{v_i\}}.$$

In the case of a Cobb-Douglas production function, $TE_i = \exp\{-u_i\}$. See the section "Stochastic Frontier Production and Cost Models" on page 1992.

Cost frontier can be written in general as

$$E_i = c(y_i, w_i;\beta)\exp\{v_i\}/CE_i$$

where w_i denotes producer i's input prices, $c(\cdot)$ is the deterministic part of cost frontier, $\exp\{v_i\}$ is a producer-specific error term, and CE_i is the cost efficiency coefficient, which can be written as

$$CE_i = \frac{c(x_i, w_i;\beta)\exp\{v_i\}}{E_i}$$

In the case of a Cobb-Douglas cost function, $CE_i = \exp\{-u_i\}$. See the section "Stochastic Frontier Production and Cost Models" on page 1992. Hence, both technical and cost efficiency coefficients are the same. The estimates of technical efficiency are provided in the following subsections.

Normal–Half Normal Model

Define $\mu_* = -\epsilon\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. Then, as it is shown by Jondrow et al. (1982), conditional density is as follows:

$$f(u|\epsilon) = \frac{f(u,\epsilon)}{f(\epsilon)} = \frac{1}{\sqrt{2\pi}\sigma_*}\exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\} \Big/ \left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]$$

Hence, $f(u|\epsilon)$ is the density for $N^+(\mu_*, \sigma_*^2)$.

Using this result, it follows that the estimate of technical efficiency (Battese and Coelli 1988) is

$$TE1_i = E(\exp\{-u_i\}|\epsilon_i) = \left[\frac{1-\Phi(\sigma_* - \mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)}\right]\exp\left\{-\mu_{*i}+\frac{1}{2}\sigma_*^2\right\}$$

The second version of the estimate (Jondrow et al. 1982) is

$$TE2_i = \exp\{-E(u_i|\epsilon_i)\}$$

where

$$E(u_i|\epsilon_i) = \mu_{*i} + \sigma_*\left[\frac{\phi(-\mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)}\right] = \sigma_*\left[\frac{\phi(\epsilon_i\lambda/\sigma)}{1-\Phi(\epsilon_i\lambda/\sigma)} - \left(\frac{\epsilon_i\lambda}{\sigma}\right)\right]$$

Normal-Exponential Model

Define $A = -\tilde{\mu}/\sigma_v$ and $\tilde{\mu} = -\epsilon - \sigma_v^2/\sigma_u$. Then, as it is shown by Kumbhakar and Lovell (2000), conditional density is as follows:

$$f(u|\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_v \Phi(-\tilde{\mu}/\sigma_v)} \exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma^2}\right\}$$

Hence, $f(u|\epsilon)$ is the density for $N^+(\tilde{\mu}, \sigma_v^2)$.

Using this result, it follows that the estimate of technical efficiency is

$$TE1_i = E(\exp\{-u_i\}|\epsilon_i) = \left[\frac{1-\Phi(\sigma_v - \tilde{\mu}_i/\sigma_v)}{1-\Phi(-\tilde{\mu}_i/\sigma_v)}\right] \exp\left\{-\tilde{\mu}_i + \frac{1}{2}\sigma_v^2\right\}$$

The second version of the estimate is

$$TE2_i = \exp\{-E(u_i|\epsilon_i)\}$$

where

$$E(u_i|\epsilon_i) = \tilde{\mu}_i + \sigma_v\left[\frac{\phi(-\tilde{\mu}_i/\sigma_v)}{1-\Phi(-\tilde{\mu}_i/\sigma_v)}\right] = \sigma_v\left[\frac{\phi(A)}{\Phi(-A)} - A\right]$$

Normal–Truncated Normal Model

Define $\tilde{\mu} = (-\sigma_u^2\epsilon_i + \mu\sigma_v^2)/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. Then, as it is shown by Kumbhakar and Lovell (2000), conditional density is as follows:

$$f(u|\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_*[1-\Phi(-\tilde{\mu}/\sigma_*)]} \exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma_*^2}\right\}$$

Hence, $f(u|\epsilon)$ is the density for $N^+(\tilde{\mu}, \sigma_*^2)$.

Using this result, it follows that the estimate of technical efficiency is

$$TE1_i = E(\exp\{-u_i\}|\epsilon_i) = \frac{1-\Phi(\sigma_* - \tilde{\mu}_i/\sigma_*)}{1-\Phi(-\tilde{\mu}_i/\sigma_*)} \exp\left\{-\tilde{\mu}_i + \frac{1}{2}\sigma_*^2\right\}$$

The second version of the estimate is

$$TE2_i = \exp\{-E(u_i|\epsilon_i)\}$$

where

$$E(u_i|\epsilon_i) = \tilde{\mu}_i + \sigma_*\left[\frac{\phi(\tilde{\mu}_i/\sigma_*)}{1-\Phi(-\tilde{\mu}_i/\sigma_*)}\right]$$

OUTEST= Data Set

The OUTEST= data set contains all the parameters estimated in a MODEL statement. The OUTEST= option can be used when the PROC QLIM call contains one MODEL statement:

```
proc qlim data=a outest=e;
   model y = x1 x2 x3;
   endogenous y ~ censored(lb=0);
run;
```

Each parameter contains the estimate for the corresponding parameter in the corresponding model. In addition, the OUTEST= data set contains the following variables:

NAME	the name of the independent variable
TYPE	type of observation. PARM indicates the row of coefficients; STD indicates the row of standard deviations of the corresponding coefficients.
STATUS	convergence status for optimization

The rest of the columns correspond to the explanatory variables.

The OUTEST= data set contains one observation for the MODEL statement, giving the parameter estimates for that model. If the COVOUT option is specified, the OUTEST= data set includes additional observations for the MODEL statement, giving the rows of the covariance matrix of parameter estimates. For covariance observations, the value of the _TYPE_ variable is COV, and the _NAME_ variable identifies the parameter associated with that row of the covariance matrix. If the CORROUT option is specified, the OUTEST= data set includes additional observations for the MODEL statement, giving the rows of the correlation matrix of parameter estimates. For correlation observations, the value of the _TYPE_ variable is CORR, and the _NAME_ variable identifies the parameter associated with that row of the correlation matrix.

Naming

Naming of Parameters

When there is only one equation in the estimation, parameters are named in the same way as in other SAS procedures such as REG, PROBIT, and so on. The constant in the regression equation is called Intercept. The coefficients on independent variables are named by the independent variables. The standard deviation of the errors is called _Sigma. If there are Box-Cox transformations, the coefficients are named _Lambda*i*, where *i* increments from 1, or as specified by the user. The limits for the discrete dependent variable are named _Limit*i*. If the LIMIT=varying option is specified, then _Limit*i* starts from 1. If the LIMIT=varying option is not specified, then _Limit1 is set to 0 and the limit parameters start from $i = 2$. If the HETERO statement is included, the coefficients of the independent variables in the hetero equation are called _H.*x*, where *x* is the name of the independent variable. You can form the name of the parameter associated with an interaction regressor by concatenating the interacting variables with an underscore. The following example restricts the parameter that includes the interaction term to be greater than zero:

```
proc qlim data=a;
   model y = x1|x2;
   endogenous y ~ discrete;
   restrict x1_x2>0;
run;
```

When there are multiple equations in the estimation, the parameters in the main equation are named in the format of *y*.*x*, where *y* is the name of the dependent variable and *x* is the name of the independent variable. The

standard deviation of the errors is called _Sigma.*y*. The correlation of the errors is called _Rho for bivariate model. For the model with three variables it is _Rho.*y1*.*y2*, _Rho.*y1*.*y3*, _Rho.*y2*.*y3*. The construction of correlation names for multivariate models is analogous. Box-Cox parameters are called _Lambda*i*.*y* and limit variables are called _Limit*i*.*y*. Parameters in the HETERO statement are named as _H.*y*.*x*. In the OUTEST= data set, all variables are changed from '.' to '_'.

Naming of Output Variables

Table 28.11 shows the option in the OUTPUT statement, with the corresponding variable names and their explanation.

Table 28.11 OUTPUT Statement Options and Variable Names

Option	Name	Explanation
PREDICTED	P_y	Predicted value of *y*
RESIDUAL	RESID_y	Residual of *y*, (y-PredictedY)
XBETA	XBETA_y	Structure part ($\mathbf{x}'\boldsymbol{\beta}$) of *y* equation
ERRSTD	ERRSTD_y	Standard deviation of error term
PROB	PROB_y	Probability that *y* is taking the observed value in this observation (discrete *y* only)
PROBALL	PROB*i*_y	Probability that *y* is taking the *i*th value (discrete *y* only)
MILLS	MILLS_y	Inverse Mills ratio for *y*
EXPECTED	EXPCT_y	Unconditional expected value of *y*
CONDITIONAL	CEXPCT_y	Conditional expected value of *y*, condition on the truncation.
MARGINAL	MEFF_x	Marginal effect of *x* on *y* ($\frac{\partial y}{\partial x}$) with single equation
	MEFF_y_x	Marginal effect of *x* on *y* ($\frac{\partial y}{\partial x}$) with multiple equations
	MEFF_P*i*_x	Marginal effect of *x* on *y* ($\frac{\partial \mathrm{Prob}(y=i)}{\partial x}$) with single equation and discrete *y*
	MEFF_P*i*_y_x	Marginal effect of *x* on *y* ($\frac{\partial \mathrm{Prob}(y=i)}{\partial x}$) with multiple equations and discrete *y*
TE1	TE1	Technical efficiency estimate for each producer proposed by Battese and Coelli (1988)
TE2	TE2	Technical efficiency estimate for each producer proposed by Jondrow et al. (1982)

If you prefer to name the output variables differently, you can use the RENAME option in the data set. For example, the following statements rename the residual of *y* as *Resid*:

```
proc qlim data=one;
   model y = x1-x10 / censored;
   output out=outds(rename=(resid_y=resid)) residual;
run;
```

ODS Table Names

PROC QLIM assigns a name to each table it creates. You can use these names to denote the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 28.12.

Table 28.12 ODS Tables Produced in PROC QLIM by the MODEL Statement and TEST Statement

ODS Table Name	Description	Option
ODS Tables Created by the MODEL Statement and TEST Statement		
ResponseProfile	Response profile	Default
ClassLevels	Class levels	Default
FitSummary	Summary of nonlinear estimation	Default
GoodnessOfFit	Pseudo-R-square measures	Default
ConvergenceStatus	Convergence status	Default
ParameterEstimates	Parameter estimates	Default
SummaryContResponse	Summary of continuous response	Default
CovB	Covariance of parameter estimates	COVB
CorrB	Correlation of parameter estimates	CORRB
FitSummaryHeckman1	Heckman First Step Model Fit Summary	HECKIT
FitSummaryHeckman2	Heckman Second Model Fit Summary	HECKIT
LinCon	Linear constraints	ITPRINT
InputOptions	Input options	ITPRINT
ProblemDescription	Problem description	ITPRINT
IterStart	Optimization start summary	ITPRINT
IterHist	Iteration history	ITPRINT
IterStop	Optimization results	ITPRINT
ConvergenceStatus	Convergence status	ITPRINT
ParameterEstimatesStart	Optimization start	ITPRINT
ParameterEstimatesResults	Resulting parameters	ITPRINT
LinConSol	Linear constraints evaluated at solution	ITPRINT
VariableSelection	Variable selection summary	SELECTVAR
ODS Tables Created by the TEST Statement		
TestResults	Test results	Default
ODS Tables Created by the BAYES Statement		
AutoMcmcSummary	Automatic MCMC summary	DIAGNOSTICS=AUTOSUM
AutoCorr	Autocorrelation statistics for each parameter	Default
Corr	Correlation matrix of the posterior samples	STATS=COR

Table 28.12 *continued*

ODS Table Name	Description	Option
Cov	Covariance matrix of the posterior samples	STATS=COV
ESS	Effective sample size for each parameter	Default
MCSE	Monte Carlo standard error for each parameter	Default
Geweke	Geweke diagnostics for each parameter	Default
Heidelberger	Heidelberger-Welch diagnostics for each parameter	DIAGNOSTICS=HEIDEL
LogMarginLike	Marginal likelihood	MARGINLIKE
PostIntervals	Equal-tail and HPD intervals for each parameter	Default
PosteriorSample	Posterior samples	(ODS output data set only)
PostSummaries	Posterior summaries	Default
PriorSample	Prior samples used for prior predictive analysis	(ODS output data set only)
PriorSummaries	Prior summaries	STATS=PRIOR
Raftery	Raftery-Lewis diagnostics for each parameter	DIAGNOSTICS=RAFTER
ODS Tables Created by the RANDOM Statement		
RandParmsModelSummary	Random-parameters model summary	Default
RandParmsCovEstimates	Random-parameters covariance estimates	Default

ODS Graphics

You can reference every graph that is produced through ODS Graphics with a name. The names of the graphs that PROC QLIM generates are listed in Table 28.13 for the frequentist approach and in Table 28.14 for the Bayesian approach.

Table 28.13 Graphs Produced by PROC QLIM without a BAYES Statement

ODS Graph Name	Plot Description	Statement and Option
Frequentist Output Plots		
ResidPlot	Frequentist analysis of residuals	PLOTS=RESIDUAL
XbetaPlot	Frequentist analysis of xbeta	PLOTS=XBETA
PredPlot	Frequentist analysis of predictions	PLOTS=PREDICTED
MarginalPlot	Frequentist analysis of marginal effects	PLOTS=MARGINAL
ErrStdPlot	Frequentist analysis of the error standard deviation (meaningful only with a HETERO statement)	PLOTS=ERRSTD
MillsPlot	Frequentist analysis of Mills ratio	PLOTS=MILLS
ExpctPlot	Frequentist analysis of expected values for continuous endogenous variables	PLOTS=EXPECTED

Table 28.13 *continued*

ODS Graph Name	Plot Description	Statement and Option
TE1Plot	Frequentist analysis of technical efficiency (only in stochastic frontier model) suggested by Battese and Coelli (1988)	PLOTS=TE1
TE2Plot	Frequentist analysis of technical efficiency (only in stochastic frontier model) suggested by Jondrow et al. (1982)	PLOTS=TE2
CExpctPlot	Frequentist analysis of conditional expected values for continuous endogenous variables	PLOTS=CONDITIONAL
ProbPlot	Frequentist analysis of probability of discrete endogenous variables that take the current observed responses	PLOTS=PROB
ProbAllPlot	Frequentist analysis of probability of discrete endogenous variables for all responses	PLOTS=PROBALL
ProfLikPlot	Profile log-likelihood plot	PLOTS=PROFLIK

Table 28.14 Graphs Produced by PROC QLIM When a BAYES Statement Is Included

ODS Graph Name	Plot Description	Statement and Option
Bayesian Diagnostic Plots		
ADPanel	Autocorrelation function and density panel	PLOTS=(AUTOCORR DENSITY)
AutocorrPanel	Autocorrelation function panel	PLOTS=AUTOCORR
AutocorrPlot	Autocorrelation function plot	PLOTS(UNPACK)=AUTOCORR
DensityPanel	Density panel	PLOTS=DENSITY
DensityPlot	Density plot	PLOTS(UNPACK)=DENSITY
ProfLikPlot	Profile log-likelihood plot	PLOTS=PROFLIK
TAPanel	Trace and autocorrelation function panel	PLOTS=(TRACE AUTOCORR)
TADPanel	Trace, density, and autocorrelation function panel	PLOTS=(TRACE AUTOCORR DENSITY) PLOTS=BAYESDIAG
TDPanel	Trace and density panel	PLOTS=(TRACE DENSITY)
TracePanel	Trace panel	PLOTS=TRACE
TracePlot	Trace plot	PLOTS(UNPACK)=TRACE
Bayesian Summary Plots		
BayesSumPlot	Prior/posterior densities and MLE	PLOTS=BAYESSUM
Bayesian Output Plots		
PredictiveByObsNumPlot	Predictive analysis by observation number	PLOTS(PRIOR)=BAYESPRED

Table 28.14 *continued*

ODS Graph Name	Plot Description	Statement and Option
PredictivePlot	Predictive analysis by regressor	PLOTS(PRIOR)=BAYESPRED

The ODS Graphics is not supported for the random-parameters models.

Examples: QLIM Procedure

Example 28.1: Ordered Data Modeling

Cameron and Trivedi (1986, 1998) studied the number of doctor visits from the Australian Health Survey, 1977–1978. In the following data set, the dependent variable, DVISITS, contains the number of doctor visits in the past 2 weeks (0, 1, or more than 2). The explanatory variables are as follows: SEX indicates if the patient is female; AGE is the age in years divided by 100; INCOME is the annual income ($10,000); LEVYPLUS indicates if the patient has private health insurance; FREEPOOR indicates free government health insurance due to low income; FREEREPA indicates free government health insurance for other reasons; ILLNESS is the number of illnesses in the past 2 weeks; ACTDAYS is the number of days the illness caused reduced activity; HSCORE is a questionnaire score; CHCOND1 indicates a chronic condition that does not limit activity; and CHCOND2 indicates a chronic condition that limits activity.

```
data docvisit;
   input sex age agesq income levyplus freepoor freerepa
         illness actdays hscore chcond1 chcond2 dvisits;
   y = (dvisits > 0);
   if ( dvisits > 8 ) then dvisits = 8;
datalines;
1 0.19 0.0361 0.55  1  0  0  1  4  1  0  0  1
1 0.19 0.0361 0.45  1  0  0  1  2  1  0  0  1

   ... more lines ...

1 0.37 0.1369 0.25  0  0  1  1  0  1  0  0  0
1 0.52 0.2704 0.65  0  0  0  0  0  0  0  0  0
0 0.72 0.5184 0.25  0  0  1  0  0  0  0  0  0
;
```

The dependent variable, DVISITS, has nine ordered values. The following SAS statements estimate the ordinal probit model:

```
/*-- Ordered Discrete Responses --*/
proc qlim data=docvisit;
   model dvisits = sex age agesq income levyplus
                   freepoor freerepa illness actdays hscore
                   chcond1 chcond2 / discrete;
run;
```

The output of the QLIM procedure for ordered data modeling is shown in Output 28.1.1.

Output 28.1.1 Ordered Data Modeling

Binary Data

The QLIM Procedure

Discrete Response Profile of dvisits

Index	Value	Total Frequency
1	0	4141
2	1	782
3	2	174
4	3	30
5	4	24
6	5	9
7	6	12
8	7	12
9	8	6

Output 28.1.1 *continued*

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	dvisits
Number of Observations	5190
Log Likelihood	-3138
Maximum Absolute Gradient	0.0003675
Number of Iterations	82
Optimization Method	Quasi-Newton
AIC	6316
Schwarz Criterion	6447

Goodness-of-Fit Measures

Measure	Value	Formula
Likelihood Ratio (R)	789.73	2 * (LogL - LogL0)
Upper Bound of R (U)	7065.9	- 2 * LogL0
Aldrich-Nelson	0.1321	R / (R+N)
Cragg-Uhler 1	0.1412	1 - exp(-R/N)
Cragg-Uhler 2	0.1898	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.149	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.1416	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.1118	R / U
Veall-Zimmermann	0.2291	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.2036	

N = # of observations, K = # of regressors

Output 28.1.1 *continued*

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-1.378705	0.147413	-9.35	<.0001
sex	1	0.131885	0.043785	3.01	0.0026
age	1	-0.534190	0.815907	-0.65	0.5126
agesq	1	0.857308	0.898364	0.95	0.3399
income	1	-0.062211	0.068017	-0.91	0.3604
levyplus	1	0.137030	0.053262	2.57	0.0101
freepoor	1	-0.346045	0.129638	-2.67	0.0076
freerepa	1	0.178382	0.074348	2.40	0.0164
illness	1	0.150485	0.015747	9.56	<.0001
actdays	1	0.100575	0.005850	17.19	<.0001
hscore	1	0.031862	0.009201	3.46	0.0005
chcond1	1	0.061601	0.049024	1.26	0.2089
chcond2	1	0.135321	0.067711	2.00	0.0457
_Limit2	1	0.938884	0.031219	30.07	<.0001
_Limit3	1	1.514288	0.049329	30.70	<.0001
_Limit4	1	1.711660	0.058151	29.43	<.0001
_Limit5	1	1.952860	0.072014	27.12	<.0001
_Limit6	1	2.087422	0.081655	25.56	<.0001
_Limit7	1	2.333786	0.101760	22.93	<.0001
_Limit8	1	2.789796	0.156189	17.86	<.0001

By default, ordinal probit/logit models are estimated assuming that the first threshold or limit parameter (μ_1) is 0. However, this parameter can also be estimated when the LIMIT1=VARYING option is specified. The probability that y_i^* belongs to the *j*th category is defined as

$$P[\mu_{j-1} < y_i^* < \mu_j] = F(\mu_j - \mathbf{x}_i'\boldsymbol{\beta}) - F(\mu_{j-1} - \mathbf{x}_i'\boldsymbol{\beta})$$

where $F(\cdot)$ is the logistic or standard normal CDF, $\mu_0 = -\infty$ and $\mu_9 = \infty$. Output 28.1.2 lists ordinal probit estimates computed in the following program. Note that the intercept term is suppressed for model identification when μ_1 is estimated.

```
/*-- Ordered Probit --*/
proc qlim data=docvisit;
   model dvisits = sex age agesq income levyplus
                   freepoor freerepa illness actdays hscore
                   chcond1 chcond2 / discrete(d=normal) limit1=varying;
run;
```

Output 28.1.2 Ordinal Probit Parameter Estimates with LIMIT1=VARYING

Binary Data

The QLIM Procedure

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
sex	1	0.131885	0.043785	3.01	0.0026
age	1	-0.534181	0.815915	-0.65	0.5127
agesq	1	0.857298	0.898371	0.95	0.3399
income	1	-0.062211	0.068017	-0.91	0.3604
levyplus	1	0.137031	0.053262	2.57	0.0101
freepoor	1	-0.346045	0.129638	-2.67	0.0076
freerepa	1	0.178382	0.074348	2.40	0.0164
illness	1	0.150485	0.015747	9.56	<.0001
actdays	1	0.100575	0.005850	17.19	<.0001
hscore	1	0.031862	0.009201	3.46	0.0005
chcond1	1	0.061602	0.049024	1.26	0.2089
chcond2	1	0.135322	0.067711	2.00	0.0457
_Limit1	1	1.378706	0.147415	9.35	<.0001
_Limit2	1	2.317590	0.150206	15.43	<.0001
_Limit3	1	2.892994	0.155198	18.64	<.0001
_Limit4	1	3.090367	0.158263	19.53	<.0001
_Limit5	1	3.331566	0.164065	20.31	<.0001
_Limit6	1	3.466128	0.168799	20.53	<.0001
_Limit7	1	3.712493	0.179756	20.65	<.0001
_Limit8	1	4.168502	0.215738	19.32	<.0001

Example 28.2: Tobit Analysis

The following statements show a subset of the Mroz (1987) data set. In these data, Hours is the number of hours the wife worked outside the household in a given year, Yrs_Ed is the years of education, and Yrs_Exp is the years of work experience. A Tobit model will be fit to the hours worked with years of education and experience as covariates.

By the nature of the data, it is clear that there are a number of women who committed some positive number of hours to outside work ($y_i > 0$ is observed). There are also a number of women who did not work at all ($y_i = 0$ is observed). This produces the model

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$$

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}$$

where $\epsilon_i \sim \text{iid} N(0, \sigma^2)$. The set of explanatory variables is denoted by $\mathbf{x}_i$.

```
title1 'Estimating a Tobit model';

data subset;
   input Hours Yrs_Ed Yrs_Exp @@;
   if Hours eq 0 then Lower=.;
      else               Lower=Hours;
datalines;
0 8 9 0 8 12 0 9 10 0 10 15 0 11 4 0 11 6
1000 12 1 1960 12 29 0 13 3 2100 13 36
3686 14 11 1920 14 38 0 15 14 1728 16 3
1568 16 19 1316 17 7 0 17 15
;

/*-- Tobit Model --*/
proc qlim data=subset;
   model hours = yrs_ed yrs_exp;
   endogenous hours ~ censored(lb=0);
run;
```

The output of the QLIM procedure is shown in Output 28.2.1.

Output 28.2.1 Tobit Analysis Results

Estimating a Tobit model

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	Hours
Number of Observations	17
Log Likelihood	-74.93700
Maximum Absolute Gradient	1.18953E-6
Number of Iterations	23
Optimization Method	Quasi-Newton
AIC	157.87400
Schwarz Criterion	161.20685

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-5598.295129	27.692220	-202.16	<.0001
Yrs_Ed	1	373.123254	53.988877	6.91	<.0001
Yrs_Exp	1	63.336247	36.551299	1.73	0.0831
_Sigma	1	1582.859635	390.076480	4.06	<.0001

The "Parameter Estimates" table has four rows. The first three of these rows correspond to the vector estimate of the regression coefficients $\boldsymbol{\beta}$. The last one is called _Sigma, which corresponds to the estimate of the error variance σ.

Example 28.3: Bivariate Probit Analysis

This example shows how to estimate a bivariate probit model. Note the INIT statement in the following program, which sets the initial values for some parameters in the optimization:

```
data a;
   keep y1 y2 x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      y1l = 1 + 2 * x1 + 3 * x2 + u1;
      y2l = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      if ( y1l > 0 ) then y1 = 1;
      else                y1 = 0;
      if ( y2l > 0 ) then y2 = 1;
      else                y2 = 0;
      output;
   end;
run;

/*-- Bivariate Probit --*/
proc qlim data=a method=qn;
   init y1.x1 2.8, y1.x2 2.1, _rho .1;
   model y1 = x1 x2;
   model y2 = x1 x2;
   endogenous y1 y2 ~ discrete;
run;
```

The output of the QLIM procedure is shown in Output 28.3.1.

Output 28.3.1 Bivariate Probit Analysis Results

Estimating a Tobit model

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	2
Endogenous Variable	y1 y2
Number of Observations	500
Log Likelihood	-134.90796
Maximum Absolute Gradient	3.23363E-7
Number of Iterations	17
Optimization Method	Quasi-Newton
AIC	283.81592
Schwarz Criterion	313.31817

Output 28.3.1 *continued*

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y1.Intercept	1	1.003639	0.153678	6.53	<.0001
y1.x1	1	2.244374	0.256062	8.76	<.0001
y1.x2	1	3.273441	0.341581	9.58	<.0001
y2.Intercept	1	3.621164	0.457173	7.92	<.0001
y2.x1	1	4.551525	0.576547	7.89	<.0001
y2.x2	1	-2.442769	0.332295	-7.35	<.0001
_Rho	1	0.144097	0.336459	0.43	0.6685

Example 28.4: Sample Selection Model

This example illustrates the use of PROC QLIM for sample selection models. The data set is the one from Mroz (1987). The goal is to estimate a wage offer function for married women, accounting for potential selection bias. Of the 753 women, the wage is observed for 428 working women. The labor force participation equation estimated in the introductory example is used for selection. The wage equation uses log wage (lwage) as the dependent variable. The explanatory variables in the wage equation are the woman's years of schooling (educ), wife's labor experience (exper), and square of experience (expersq). The program is as follows:

```
/*-- Sample Selection --*/
proc qlim data=mroz;
   model inlf = nwifeinc educ exper expersq
                age kidslt6 kidsge6 /discrete;
   model lwage = educ exper expersq / select(inlf=1);
run;
```

The output of the QLIM procedure is shown in Output 28.4.1.

Output 28.4.1 Sample Selection

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	2
Endogenous Variable	inlf lwage
Number of Observations	753
Log Likelihood	-832.88509
Maximum Absolute Gradient	0.00502
Number of Iterations	78
Optimization Method	Quasi-Newton
AIC	1694
Schwarz Criterion	1759

Output 28.4.1 *continued*

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
lwage.Intercept	1	-0.552716	0.260371	-2.12	0.0338
lwage.educ	1	0.108351	0.014861	7.29	<.0001
lwage.exper	1	0.042837	0.014878	2.88	0.0040
lwage.expersq	1	-0.000837	0.000417	-2.01	0.0449
_Sigma.lwage	1	0.663397	0.022706	29.22	<.0001
inlf.Intercept	1	0.266459	0.508954	0.52	0.6006
inlf.nwifeinc	1	-0.012132	0.004877	-2.49	0.0129
inlf.educ	1	0.131341	0.025383	5.17	<.0001
inlf.exper	1	0.123282	0.018728	6.58	<.0001
inlf.expersq	1	-0.001886	0.000601	-3.14	0.0017
inlf.age	1	-0.052829	0.008479	-6.23	<.0001
inlf.kidslt6	1	-0.867398	0.118647	-7.31	<.0001
inlf.kidsge6	1	0.035872	0.043476	0.83	0.4093
_Rho	1	0.026617	0.147073	0.18	0.8564

Note the correlation estimate is insignificant. This indicates that selection bias is not a big problem in the estimation of wage equation.

Example 28.5: Sample Selection Model with Truncation and Censoring

In this example the data are generated such that the selection variable is discrete and the variable Y is truncated from below by zero. The program follows, with the results shown in Output 28.5.1:

```
data trunc;
   keep z y x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      zl = 1 + 2 * x1 + 3 * x2 + u1;
      y = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      if ( zl > 0 ) then z = 1;
      else z = 0;
      if y>=0 then output;
   end;
run;

/*-- Sample Selection with Truncation --*/
proc qlim data=trunc method=qn;
   model z = x1 x2 / discrete;
   model y = x1 x2 / select(z=1) truncated(lb=0);
run;
```

Output 28.5.1 Sample Selection with Truncation

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	2
Endogenous Variable	z y
Number of Observations	379
Log Likelihood	-344.10722
Maximum Absolute Gradient	4.95535E-6
Number of Iterations	17
Optimization Method	Quasi-Newton
AIC	704.21444
Schwarz Criterion	735.71473

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y.Intercept	1	3.014158	0.128548	23.45	<.0001
y.x1	1	3.995671	0.099599	40.12	<.0001
y.x2	1	-1.972697	0.096385	-20.47	<.0001
_Sigma.y	1	0.923428	0.047233	19.55	<.0001
z.Intercept	1	0.949444	0.190265	4.99	<.0001
z.x1	1	2.163928	0.288384	7.50	<.0001
z.x2	1	3.134213	0.379251	8.26	<.0001
_Rho	1	0.494356	0.176542	2.80	0.0051

In the following statements the data are generated such that the selection variable is discrete and the variable Y is censored from below by zero. The results are shown in Output 28.5.2.

```
data cens;
   keep z y x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      zl = 1 + 2 * x1 + 3 * x2 + u1;
      yl = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      if ( zl > 0 ) then z = 1;
      else               z = 0;
      if ( yl > 0 ) then y = yl;
      else               y = 0;
      output;
   end;
run;

/*-- Sample Selection with Censoring --*/
proc qlim data=cens method=qn;
   model z = x1 x2 / discrete;
   model y = x1 x2 / select(z=1) censored(lb=0);
run;
```

Output 28.5.2 Sample Selection with Censoring

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	2
Endogenous Variable	z y
Number of Observations	500
Log Likelihood	-399.78508
Maximum Absolute Gradient	2.30443E-6
Number of Iterations	19
Optimization Method	Quasi-Newton
AIC	815.57015
Schwarz Criterion	849.28702

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y.Intercept	1	3.074276	0.111617	27.54	<.0001
y.x1	1	3.963619	0.085796	46.20	<.0001
y.x2	1	-2.023548	0.088714	-22.81	<.0001
_Sigma.y	1	0.920860	0.043278	21.28	<.0001
z.Intercept	1	1.013610	0.154081	6.58	<.0001
z.x1	1	2.256922	0.255999	8.82	<.0001
z.x2	1	3.302692	0.342168	9.65	<.0001
_Rho	1	0.350776	0.197093	1.78	0.0751

Example 28.6: Types of Tobit Models

The following five examples show how to estimate different types of Tobit models (see the section "Types of Tobit Models" on page 1990). Output 28.6.1 through Output 28.6.5 show the results of the corresponding programs.

Type 1 Tobit

```
data a1;
   keep y x;
   do i = 1 to 500;
      x = rannor( 19283 );
      u = rannor( 19283 );
      yl = 1 + 2 * x + u;
      if ( yl > 0 ) then y = yl;
      else               y = 0;
      output;
   end;
run;
```

```
/*-- Type 1 Tobit --*/
proc qlim data=a1 method=qn;
   model y = x;
   endogenous y ~ censored(lb=0);
run;
```

Output 28.6.1 Type 1 Tobit

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	y
Number of Observations	500
Log Likelihood	-554.17696
Maximum Absolute Gradient	4.65556E-7
Number of Iterations	9
Optimization Method	Quasi-Newton
AIC	1114
Schwarz Criterion	1127

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.942734	0.056784	16.60	<.0001
x	1	2.049571	0.066979	30.60	<.0001
_Sigma	1	1.016571	0.039035	26.04	<.0001

Type 2 Tobit

```
data a2;
   keep y1 y2 x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      y1l = 1 + 2 * x1 + 3 * x2 + u1;
      y2l = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      if ( y1l > 0 ) then y1 = 1;
      else                y1 = 0;
      if ( y1l > 0 ) then y2 = y2l;
      else                y2 = 0;
      output;
   end;
run;
```

```
/*-- Type 2 Tobit --*/
proc qlim data=a2 method=qn;
   model y1 = x1 x2 / discrete;
   model y2 = x1 x2 / select(y1=1);
run;
```

Output 28.6.2 Type 2 Tobit

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	2
Endogenous Variable	y1 y2
Number of Observations	500
Log Likelihood	-476.12328
Maximum Absolute Gradient	8.30075E-7
Number of Iterations	17
Optimization Method	Quasi-Newton
AIC	968.24655
Schwarz Criterion	1002

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y2.Intercept	1	3.066992	0.106903	28.69	<.0001
y2.x1	1	4.004874	0.072043	55.59	<.0001
y2.x2	1	-2.079352	0.087544	-23.75	<.0001
_Sigma.y2	1	0.940559	0.039321	23.92	<.0001
y1.Intercept	1	1.017140	0.154975	6.56	<.0001
y1.x1	1	2.253080	0.256097	8.80	<.0001
y1.x2	1	3.305140	0.343695	9.62	<.0001
_Rho	1	0.292992	0.210073	1.39	0.1631

Type 3 Tobit

```
data a3;
   keep y1 y2 x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      y1l = 1 + 2 * x1 + 3 * x2 + u1;
      y2l = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      if ( y1l > 0 ) then y1 = y1l;
      else                y1 = 0;
      if ( y1l > 0 ) then y2 = y2l;
      else                y2 = 0;
      output;
   end;
```

```
run;

/*-- Type 3 Tobit --*/
proc qlim data=a3 method=qn;
   model y1 = x1 x2 / censored(lb=0);
   model y2 = x1 x2 / select(y1>0);
run;
```

Output 28.6.3 Type 3 Tobit

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	2
Endogenous Variable	y1 y2
Number of Observations	500
Log Likelihood	-838.94087
Maximum Absolute Gradient	9.71691E-6
Number of Iterations	16
Optimization Method	Quasi-Newton
AIC	1696
Schwarz Criterion	1734

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y2.Intercept	1	3.081206	0.080121	38.46	<.0001
y2.x1	1	3.998361	0.063734	62.73	<.0001
y2.x2	1	-2.088280	0.072876	-28.66	<.0001
_Sigma.y2	1	0.939799	0.039047	24.07	<.0001
y1.Intercept	1	0.981975	0.067351	14.58	<.0001
y1.x1	1	2.032675	0.059363	34.24	<.0001
y1.x2	1	2.976609	0.065584	45.39	<.0001
_Sigma.y1	1	0.969968	0.039795	24.37	<.0001
_Rho	1	0.226281	0.057672	3.92	<.0001

Type 4 Tobit

```
data a4;
   keep y1 y2 y3 x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      u3 = rannor( 19283 );
      y1l = 1 + 2 * x1 + 3 * x2 + u1;
      y2l = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      y3l = 0 - 1 * x1 + 1 * x2 + u1*.1 - u2*.5 + u3*.5;
      if ( y1l > 0 ) then y1 = y1l;
      else                y1 = 0;
      if ( y1l > 0 ) then y2 = y2l;
      else                y2 = 0;
```

```
      if ( y1l <= 0 ) then y3 = y3l;
      else                 y3 = 0;
      output;
   end;
run;

/*-- Type 4 Tobit --*/
proc qlim data=a4 method=qn;
   model y1 = x1 x2 / censored(lb=0);
   model y2 = x1 x2 / select(y1>0);
   model y3 = x1 x2 / select(y1<=0);
run;
```

Output 28.6.4 Type 4 Tobit

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	3
Endogenous Variable	y1 y2 y3
Number of Observations	500
Log Likelihood	-1128
Maximum Absolute Gradient	0.0000161
Number of Iterations	21
Optimization Method	Quasi-Newton
AIC	2285
Schwarz Criterion	2344

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y2.Intercept	1	2.894656	0.076079	38.05	<.0001
y2.x1	1	4.072704	0.062675	64.98	<.0001
y2.x2	1	-1.901163	0.076874	-24.73	<.0001
_Sigma.y2	1	0.981655	0.039564	24.81	<.0001
y3.Intercept	1	0.064594	0.179441	0.36	0.7189
y3.x1	1	-0.938384	0.096570	-9.72	<.0001
y3.x2	1	1.035798	0.123104	8.41	<.0001
_Sigma.y3	1	0.743124	0.038240	19.43	<.0001
y1.Intercept	1	0.987370	0.067861	14.55	<.0001
y1.x1	1	2.050408	0.060819	33.71	<.0001
y1.x2	1	2.982190	0.072552	41.10	<.0001
_Sigma.y1	1	1.032473	0.040971	25.20	<.0001
_Rho.y1.y2	1	0.291587	0.053436	5.46	<.0001
_Rho.y1.y3	1	-0.031665	0.260057	-0.12	0.9031

Type 5 Tobit

```
data a5;
   keep y1 y2 y3 x1 x2;
   do i = 1 to 500;
      x1 = rannor( 19283 );
      x2 = rannor( 19283 );
      u1 = rannor( 19283 );
      u2 = rannor( 19283 );
      u3 = rannor( 19283 );
      y1l = 1 + 2 * x1 + 3 * x2 + u1;
      y2l = 3 + 4 * x1 - 2 * x2 + u1*.2 + u2;
      y3l = 0 - 1 * x1 + 1 * x2 + u1*.1 - u2*.5 + u3*.5;
      if ( y1l > 0 ) then y1 = 1;
      else               y1 = 0;
      if ( y1l > 0 ) then y2 = y2l;
      else               y2 = 0;
      if ( y1l <= 0 ) then y3 = y3l;
      else               y3 = 0;
      output;
   end;
run;

/*-- Type 5 Tobit --*/
proc qlim data=a5 method=qn;
   model y1 = x1 x2 / discrete;
   model y2 = x1 x2 / select(y1>0);
   model y3 = x1 x2 / select(y1<=0);
run;
```

Output 28.6.5 Type 5 Tobit

Binary Data

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	3
Endogenous Variable	y1 y2 y3
Number of Observations	500
Log Likelihood	-734.50612
Maximum Absolute Gradient	3.57134E-7
Number of Iterations	20
Optimization Method	Quasi-Newton
AIC	1495
Schwarz Criterion	1550

Output 28.6.5 *continued*

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
y2.Intercept	1	2.887523	0.095193	30.33	<.0001
y2.x1	1	4.078926	0.069623	58.59	<.0001
y2.x2	1	-1.898898	0.086578	-21.93	<.0001
_Sigma.y2	1	0.983059	0.039987	24.58	<.0001
y3.Intercept	1	0.071764	0.171522	0.42	0.6757
y3.x1	1	-0.935299	0.092843	-10.07	<.0001
y3.x2	1	1.039954	0.120697	8.62	<.0001
_Sigma.y3	1	0.743083	0.038225	19.44	<.0001
y1.Intercept	1	1.067578	0.142789	7.48	<.0001
y1.x1	1	2.068376	0.226020	9.15	<.0001
y1.x2	1	3.157385	0.314743	10.03	<.0001
_Rho.y1.y2	1	0.312369	0.177010	1.76	0.0776
_Rho.y1.y3	1	-0.018225	0.234886	-0.08	0.9382

Example 28.7: Stochastic Frontier Models

This example illustrates the estimation of stochastic frontier production and cost models.

First, a production function model is estimated. The data for this example were collected by Christensen Associates; they represent a sample of 125 observations on inputs and output for 10 airlines between 1970 and 1984. The explanatory variables (inputs) are fuel (LF), materials (LM), equipment (LE), labor (LL), and property (LP), and (LQ) is an index that represents passengers, charter, mail, and freight transported.

The following statements create the data set:

```
title1 'Stochastic Frontier Production Model';
data airlines;
   input TS FIRM NI LQ LF LM LE LL LP;
datalines;
1 1 15 -0.0484 0.2473 0.2335 0.2294 0.2246 0.2124
1 1 15 -0.0133 0.2603 0.2492 0.241 0.2216 0.1069
2 1 15 0.088 0.2666 0.3273 0.3365 0.2039 0.0865
3 1 15 0.1619 0.3019 0.4573 0.3532 0.2346 0.0242

   ... more lines ...
```

The following statements estimate a stochastic frontier exponential production model that uses Christensen Associates data:

```
/*-- Stochastic Frontier Production Model --*/
proc qlim data=airlines;
   model LQ=LF LM LE LL LP;
   endogenous LQ ~ frontier (type=exponential production);
run;
```

Figure 28.7.1 shows the results from this production model.

Output 28.7.1 Stochastic Frontier Production Model

Stochastic Frontier Production Model

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	LQ
Number of Observations	125
Log Likelihood	83.27815
Maximum Absolute Gradient	9.92882E-7
Number of Iterations	25
Optimization Method	Quasi-Newton
AIC	-150.55630
Schwarz Criterion	-127.92979
Sigma	0.12445
Lambda	0.55766

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-0.085048	0.024528	-3.47	0.0005
LF	1	-0.115802	0.124178	-0.93	0.3511
LM	1	0.756253	0.078755	9.60	<.0001
LE	1	0.424916	0.081893	5.19	<.0001
LL	1	-0.136421	0.089702	-1.52	0.1283
LP	1	0.098967	0.042776	2.31	0.0207
_Sigma_v	1	0.108688	0.010063	10.80	<.0001
_Sigma_u	1	0.060611	0.017603	3.44	0.0006

Similarly, the stochastic frontier production function can be estimated with TYPE=HALF or TYPE=TRUNCATED options that represent half-normal and truncated normal production models.

In the next step, stochastic frontier cost function is estimated. The data for the cost model are provided by Christensen and Greene (1976). The data describe costs and production inputs of 145 U.S. electricity producers in 1955. The model being estimated follows the nonhomogeneous version of the Cobb-Douglas cost function:

$$\log\left(\frac{\text{Cost}}{\text{FPrice}}\right) = \beta_0 + \beta_1 \log\left(\frac{\text{KPrice}}{\text{FPrice}}\right) + \beta_2 \log\left(\frac{\text{LPrice}}{\text{FPrice}}\right) + \beta_3 \log(\text{Output}) + \beta_4 \frac{1}{2} \log(\text{Output})^2 + \epsilon$$

All dollar values are normalized by fuel price. The quadratic log of the output is added to capture nonlinearities due to scale effects in cost functions. New variables, log_C_PF, log_PK_PF, log_PL_PF, log_y, and log_y_sq, are created to reflect transformations. The following statements create the data set and transformed variables:

```
title1 'Stochastic Frontier Cost Model';
data electricity;
   input Firm Year Cost Output LPrice LShare KPrice KShare FPrice FShare;
datalines;
 1  1955  .0820  2.0  2.090  .3164  183.000  .4521  17.9000  .2315
 2  1955  .6610  3.0  2.050  .2073  174.000  .6676  35.1000  .1251
 3  1955  .9900  4.0  2.050  .2349  171.000  .5799  35.1000  .1852
 4  1955  .3150  4.0  1.830  .1152  166.000  .7857  32.2000  .0990

   ... more lines ...


/* Data transformations */
data electricity;
   set electricity;
   label Firm="firm index"
         Year="1955 for all observations"
         Cost="Total cost"
         Output="Total output"
         LPrice="Wage rate"
         LShare="Cost share for labor"
         KPrice="Capital price index"
         KShare="Cost share for capital"
         FPrice="Fuel price"
         FShare"Cost share for fuel";
   log_C_PF=log(Cost/FPrice);
   log_PK_PF=log(KPrice/FPrice);
   log_PL_PF=log(LPrice/FPrice);
   log_y=log(Output);
   log_y_sq=log_y**2/2;
run;
```

The following statements estimate a stochastic frontier exponential cost model that uses Christensen and Greene (1976) data:

```
/*-- Stochastic Frontier Cost Model --*/
proc qlim data=electricity;
   model log_C_PF = log_PK_PF log_PL_PF log_y log_y_sq;
   endogenous log_C_PF ~ frontier (type=exponential cost);
run;
```

Output 28.7.2 shows the results.

Output 28.7.2 Exponential Distribution

Stochastic Frontier Cost Model

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	log_C_PF
Number of Observations	159
Log Likelihood	-23.30430
Maximum Absolute Gradient	3.0458E-6
Number of Iterations	21
Optimization Method	Quasi-Newton
AIC	60.60860
Schwarz Criterion	82.09093
Sigma	0.30750
Lambda	1.71345

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-4.983211	0.543328	-9.17	<.0001
log_PK_PF	1	0.090242	0.109202	0.83	0.4086
log_PL_PF	1	0.504299	0.118263	4.26	<.0001
log_y	1	0.427182	0.066680	6.41	<.0001
log_y_sq	1	0.066120	0.010079	6.56	<.0001
_Sigma_v	1	0.154998	0.020271	7.65	<.0001
_Sigma_u	1	0.265581	0.033614	7.90	<.0001

Similarly, the stochastic frontier cost model can be estimated with TYPE=HALF or TYPE=TRUNCATED options that represent half-normal and truncated normal errors.

The following statements illustrate the half-normal option:

```
/*-- Stochastic Frontier Cost Model --*/
proc qlim data=electricity;
   model log_C_PF = log_PK_PF log_PL_PF log_y log_y_sq;
   endogenous log_C_PF ~ frontier (type=half cost);
run;
```

Output 28.7.3 shows the result.

Output 28.7.3 Half-Normal Distribution

Stochastic Frontier Cost Model

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	log_C_PF
Number of Observations	159
Log Likelihood	-34.95304
Maximum Absolute Gradient	0.0001150
Number of Iterations	22
Optimization Method	Quasi-Newton
AIC	83.90607
Schwarz Criterion	105.38840
Sigma	0.42761
Lambda	1.80031

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-4.434634	0.690197	-6.43	<.0001
log_PK_PF	1	0.069624	0.136250	0.51	0.6093
log_PL_PF	1	0.474578	0.146812	3.23	0.0012
log_y	1	0.256874	0.080777	3.18	0.0015
log_y_sq	1	0.088051	0.011817	7.45	<.0001
_Sigma_v	1	0.207637	0.039222	5.29	<.0001
_Sigma_u	1	0.373810	0.073605	5.08	<.0001

The following statements illustrate the truncated normal option:

```
/*-- Stochastic Frontier Cost Model --*/
proc qlim data=electricity;
   model log_C_PF = log_PK_PF log_PL_PF log_y log_y_sq;
   endogenous log_C_PF ~ frontier (type=truncated cost);
run;
```

Output 28.7.4 shows the results.

Output 28.7.4 Truncated Normal Distribution

Stochastic Frontier Cost Model

The QLIM Procedure

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	log_C_PF
Number of Observations	159
Log Likelihood	-36.87279
Maximum Absolute Gradient	271.78546
Number of Iterations	9
Optimization Method	Quasi-Newton
AIC	89.74557
Schwarz Criterion	114.29681
Sigma	0.30309
Lambda	1.04294E-7

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-3.772132	0.338518	-11.14	<.0001
log_PK_PF	1	-0.030841	0.143593	-0.21	0.8299
log_PL_PF	1	0.574626	0.155390	3.70	0.0002
log_y	1	0.133254	0.058093	2.29	0.0218
log_y_sq	1	0.103028	0.009912	10.39	<.0001
_Sigma_v	1	0.303087	0.016898	17.94	<.0001
_Sigma_u	1	3.1610136E-8	.	.	.
_Mu	1	0.531720	0.338538	1.57	0.1163

If no PRODUCTION or COST option is specified, the stochastic frontier production model is estimated by default.

Example 28.8: Bayesian Modeling

This example illustrates how to use the QLIM procedure to perform Bayesian analysis. The generated data mimic a hypothetical scenario in which you study the number of tickets sold for a sports event given the probability of the hosting team winning and the price of the tickets. The following statements create the data set:

```
title1 'Bayesian Analysis';

ods graphics on;

data test;
   do i=1 to 200;
      e1 = rannor(8726)*2000;
      WinChance = ranuni(8772);
```

```
      Price = 10+ranexp(8773)*4;
      y = 48000 + 5000*WinChance - 100 * price +  e1;
      if y>50000 then TicketSales = 50000;
      if y<=50000 then TicketSales = y;
      output;
   end;
   keep WinChance price  y TicketSales;
run;
```

The following statements perform Bayesian analysis of a Tobit model:

```
proc qlim data=test plots(prior)=all;
   model TicketSales = WinChance price;
   endogenous TicketSales ~ censored(lb=0 ub= 50000);
   prior intercept~normal(mean=48000);
   prior WinChance~normal(mean=5000);
   prior Price~normal(mean=-100);
   bayes NBI=10000 NMC=30000 THIN=1 ntrds=1 DIAG=ALL STATS=ALL seed=2;
run;
```

Output 28.8.1 shows the results from the maximum likelihood estimation and the Bayesian analysis with diffuse prior of this Tobit model.

Output 28.8.1 Bayesian Tobit Model

Bayesian Analysis

The QLIM Procedure

Parameter Estimates

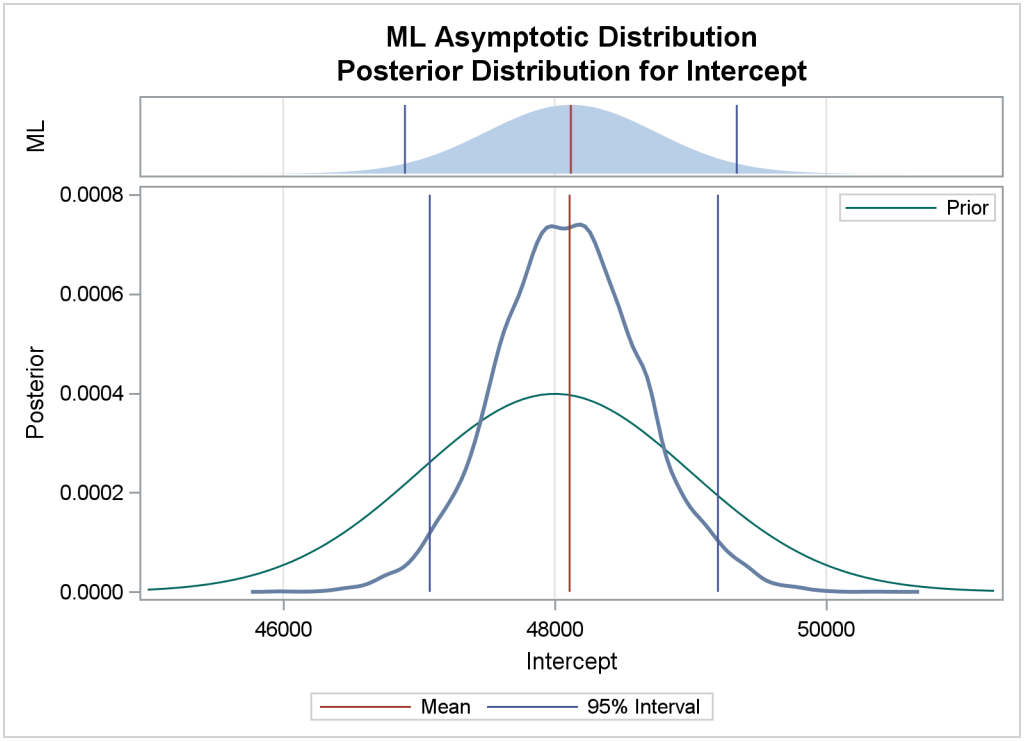
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	48119	623.565045	77.17	<.0001
WinChance	1	5242.083501	559.151222	9.38	<.0001
Price	1	-106.731665	40.660795	-2.62	0.0087
_Sigma	1	1939.607206	134.348772	14.44	<.0001

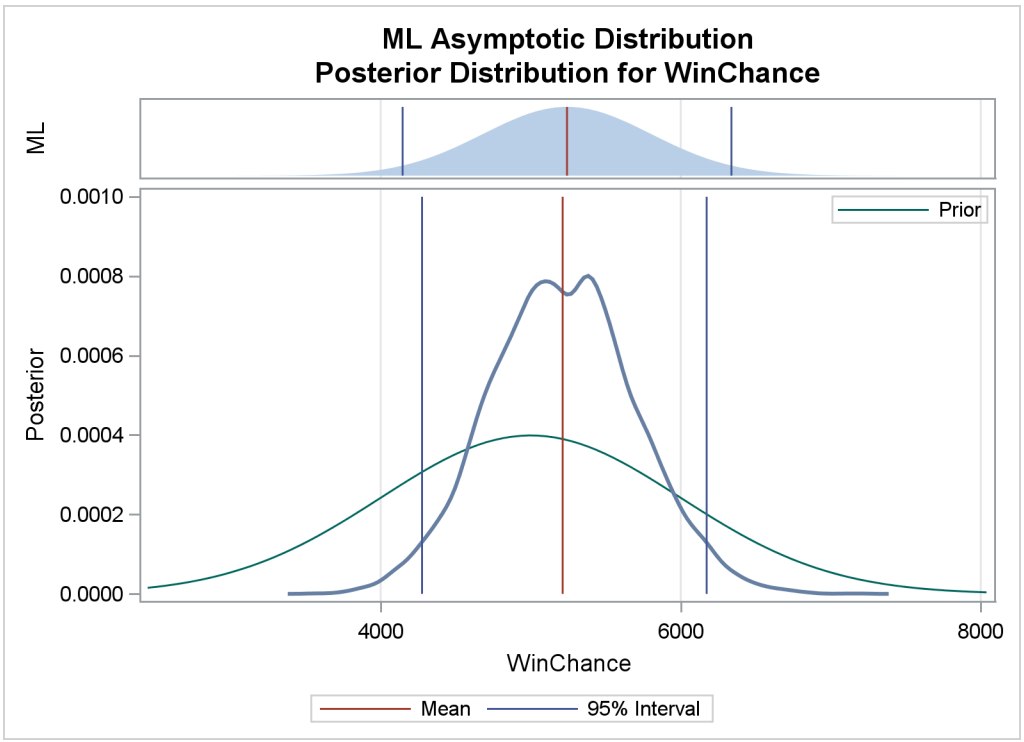
Posterior Summaries

Parameter	N	Mean	Standard Deviation	Percentiles 25%	50%	75%
Intercept	30000	48109.4	535.0	47750.5	48102.6	48460.1
WinChance	30000	5212.9	483.4	4878.8	5205.2	5533.0
Price	30000	-104.7	36.5224	-128.6	-104.2	-79.4191
_Sigma	30000	1950.9	132.9	1858.4	1945.0	2034.0

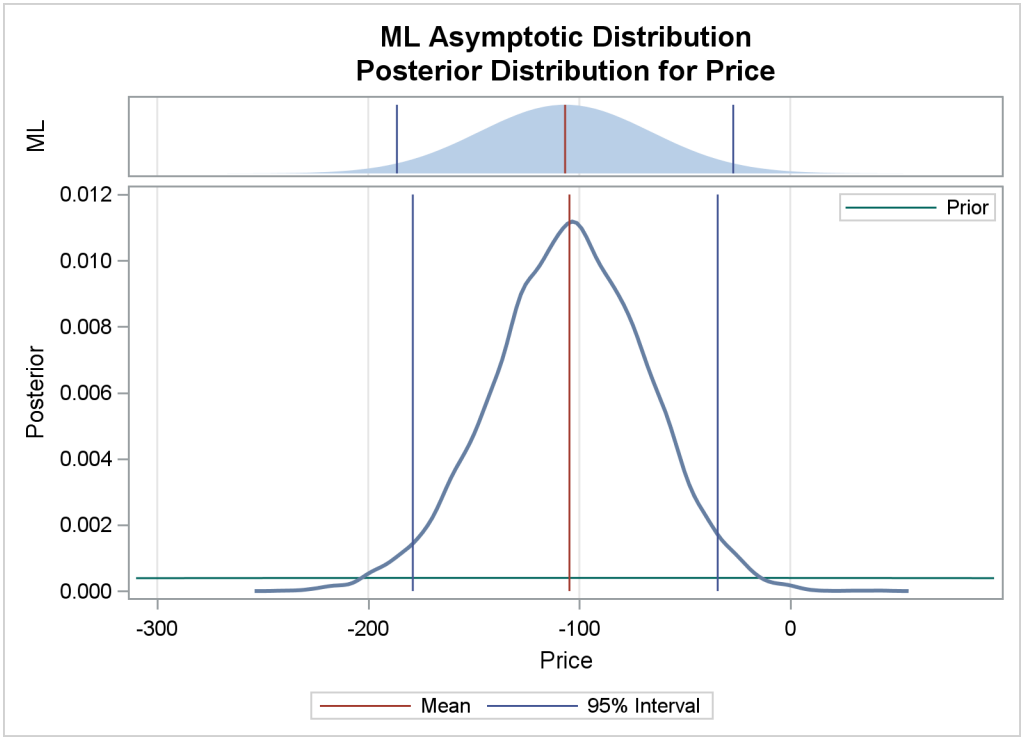
Output 28.8.2 shows a graphical representation of MLE, prior, and posterior distributions.

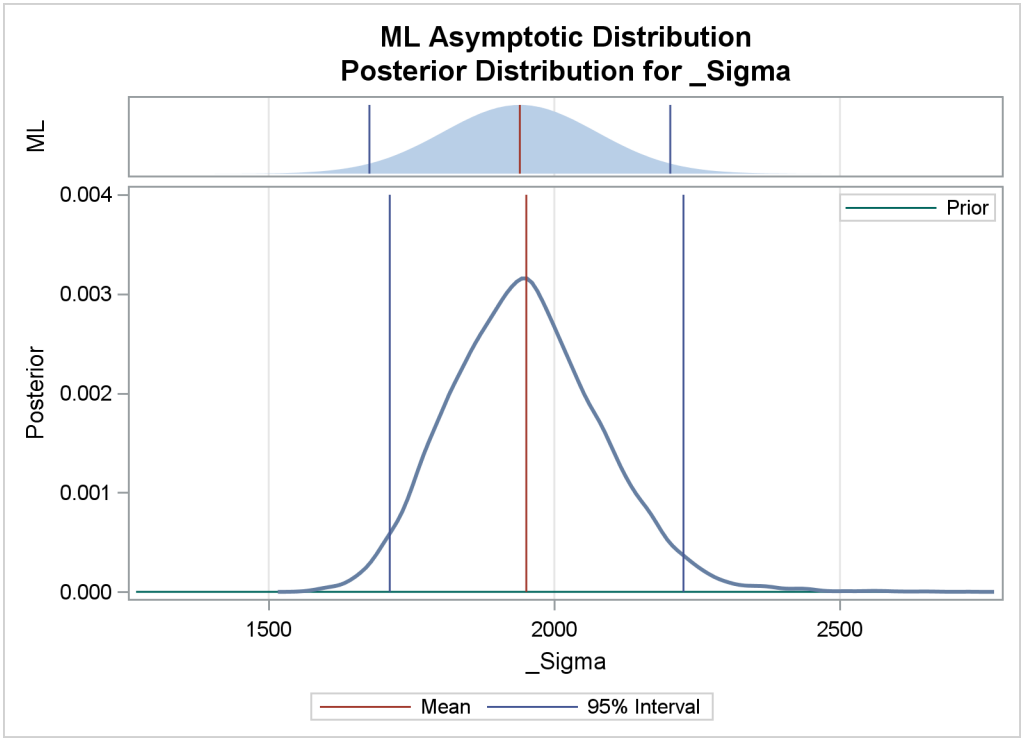
Output 28.8.2 Predictive Analysis by Observation Number





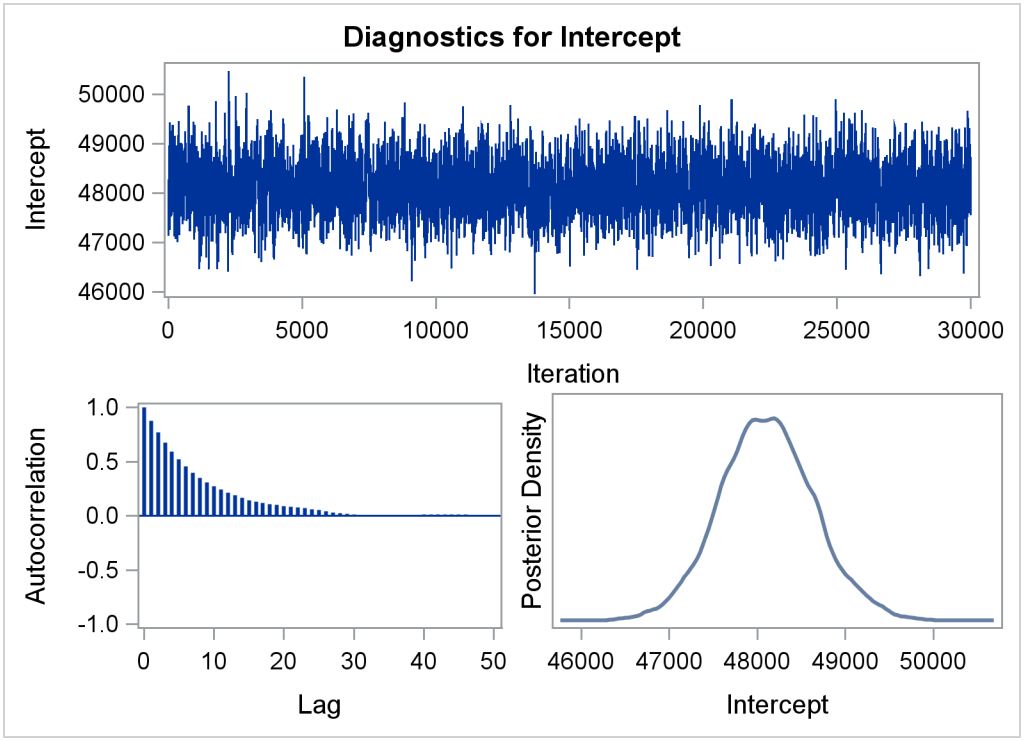
Output 28.8.2 *continued*

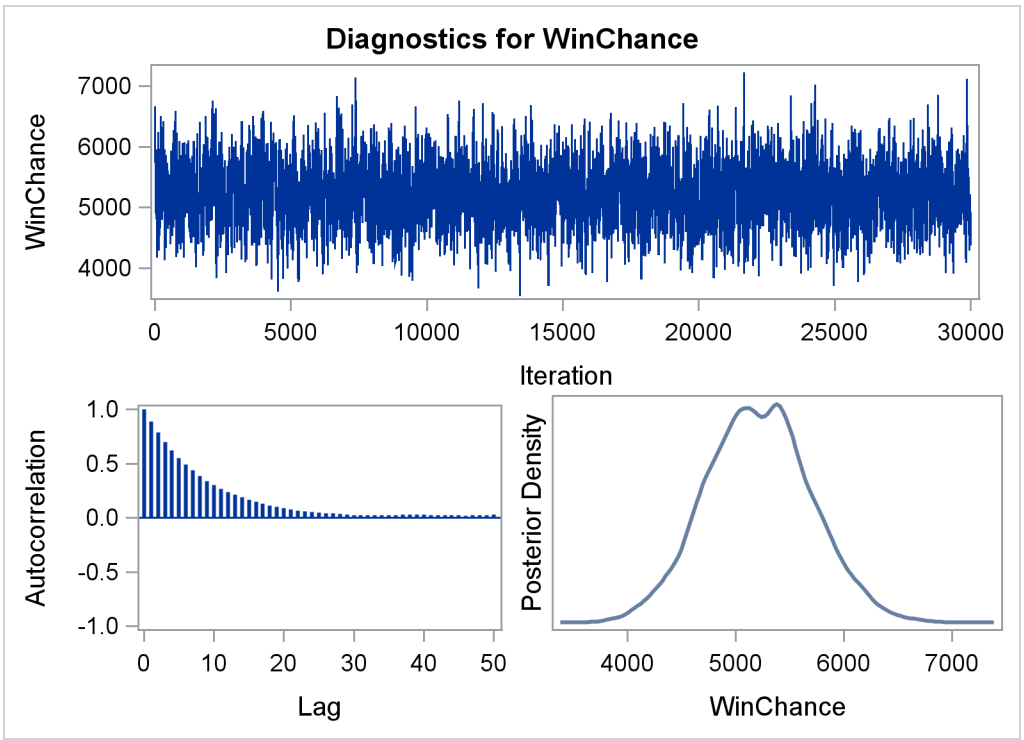




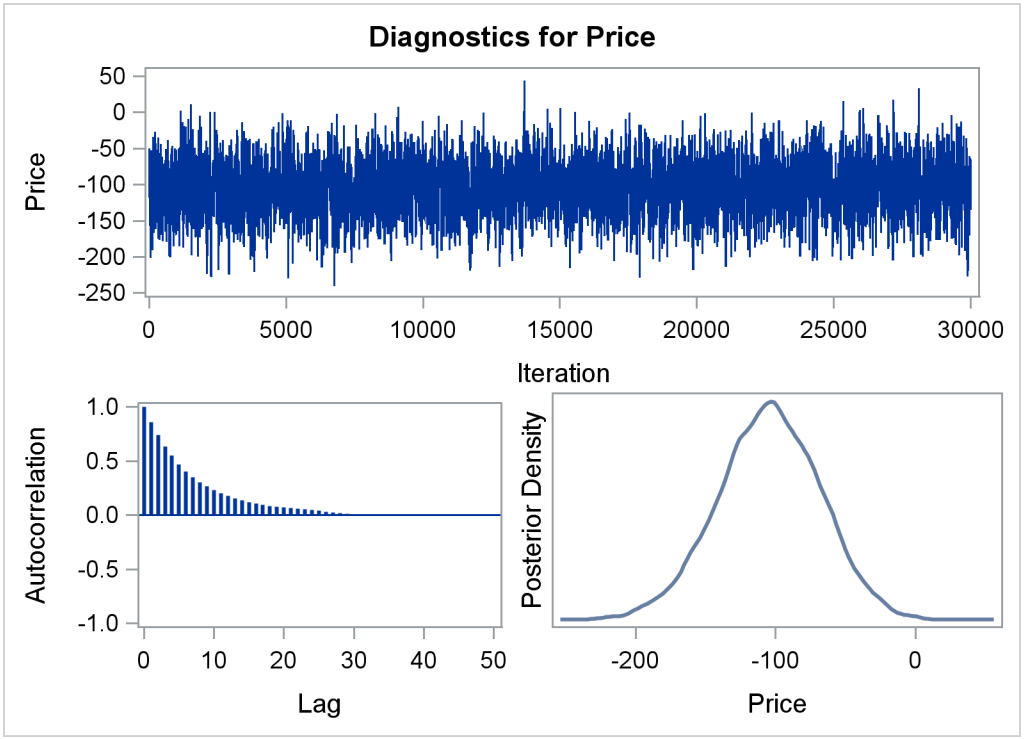
The validity of the MCMC sampling phase can be monitored with Output 28.8.3.

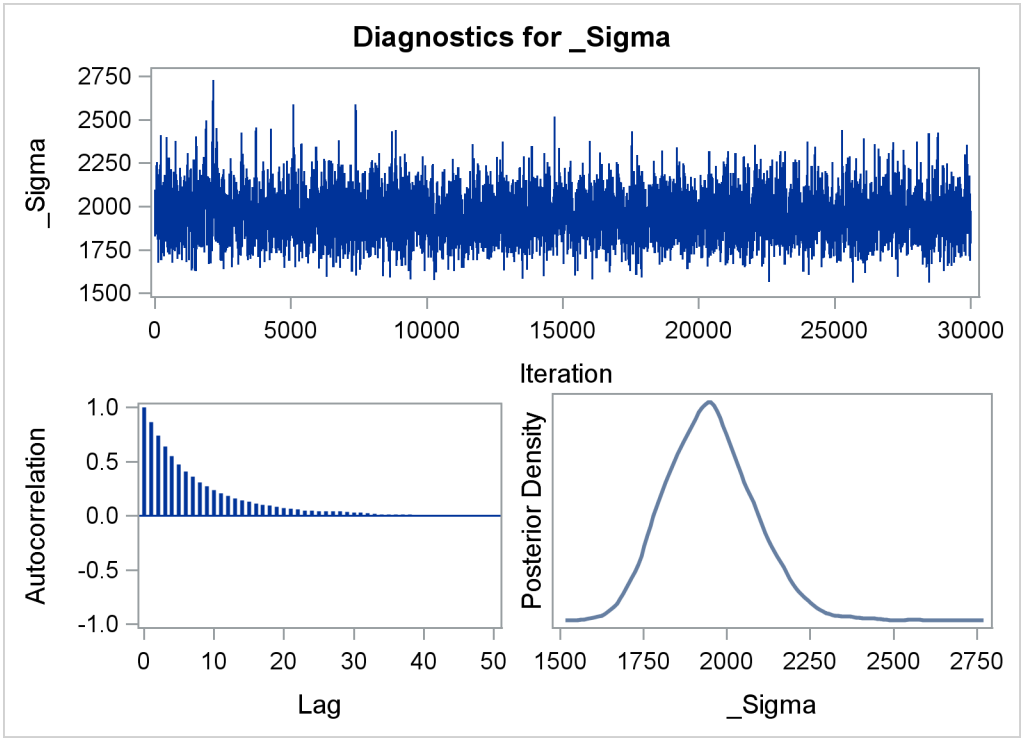
Output 28.8.3 Predictive Analysis by Observation Number





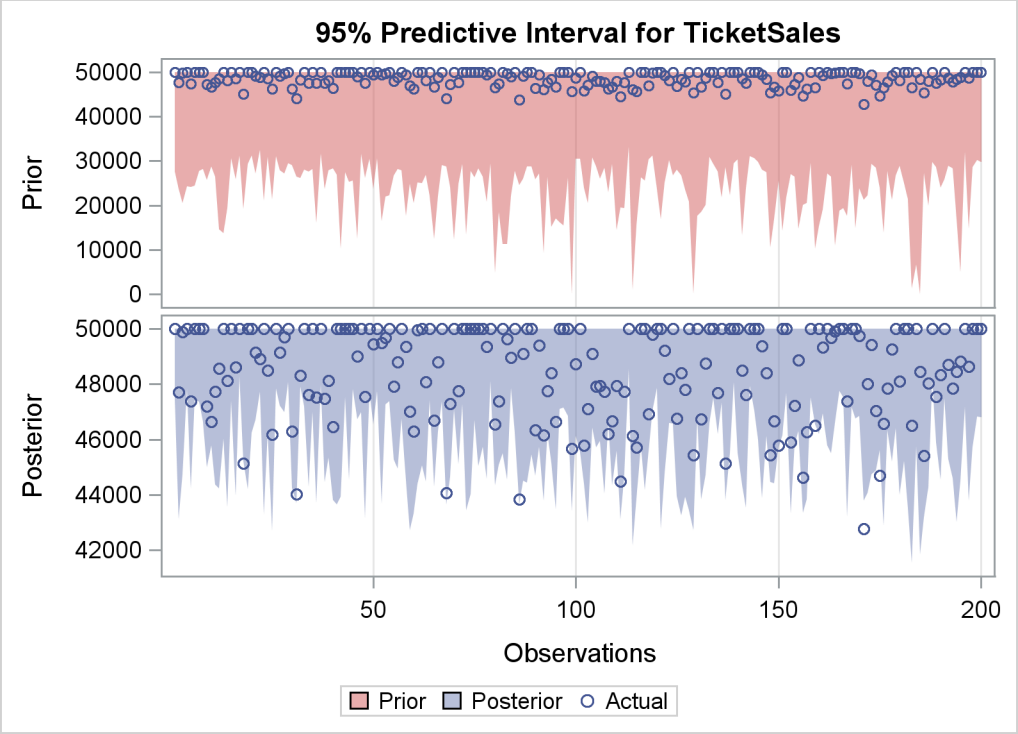
Output 28.8.3 *continued*





Finally the prior and the posterior predictive analyses are represented in Output 28.8.4.

Output 28.8.4 Predictive Analysis by Observation Number



References

Abramowitz, M., and Stegun, I. A., eds. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. 9th printing. New York: Dover.

Aigner, C., Lovell, C. A. K., and Schmidt, P. (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6:21–37.

Aitchison, J., and Silvey, S. (1957). "The Generalization of Probit Analysis to the Case of Multiple Responses." *Biometrika* 44:131–140.

Amemiya, T. (1978a). "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46:1193–1205.

Amemiya, T. (1978b). "On a Two-Step Estimate of a Multivariate Logit Model." *Journal of Econometrics* 8:13–21.

Amemiya, T. (1981). "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19:483–536.

Amemiya, T. (1984). "Tobit Models: A Survey." *Journal of Econometrics* 24:3–61.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Battese, G. E., and Coelli, T. J. (1988). "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38:387–399.

Ben-Akiva, M., and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Bera, A. K., Jarque, C. M., and Lee, L.-F. (1984). "Testing the Normality Assumption in Limited Dependent Variable Models." *International Economic Review* 25:563–578.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.

Bhat, C. R. (2001). "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." *Transportation Research, Part B* 35:677–693.

Bloom, D. E., and Killingsworth, M. R. (1985). "Correcting for Truncation Bias Caused by a Latent Truncation Variable." *Journal of Econometrics* 27:131–135.

Box, G. E. P., and Cox, D. R. (1964). "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B* 26:211–234.

Butler, J. S., and Moffitt, R. (1982). "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model." *Econometrica* 50:761–764.

Cameron, A. C., and Trivedi, P. K. (1986). "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1:29–53.

Cameron, A. C., and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Chan, J. C. C., and Eisenstat, E. (2015). "Marginal Likelihood Estimation with the Cross-Entropy Method." *Econometric Reviews* 34:256–285.

Christensen, L. R., and Greene, W. H. (1976). "Economies of Scale in U.S. Electric Power Generation." *Journal of Political Economy* 84:655–676.

Coelli, T. J., Prasada Rao, D. S., and Battese, G. E. (1998). *An Introduction to Efficiency and Productivity Analysis*. London: Kluwer Academic.

Copley, P. A., Doucet, M. S., and Gaver, K. M. (1994). "A Simultaneous Equations Analysis of Quality Control Review Outcomes and Engagement Fees for Audits of Recipients of Federal Financial Assistance." *Accounting Review* 69:244–256.

Cox, D. R. (1970). *Analysis of Binary Data*. London: Metheun.

Cox, D. R. (1972). "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Series B* 20:187–220. With discussion.

Cox, D. R. (1975). "Partial Likelihood." *Biometrika* 62:269–276.

Deis, D. R., and Hill, R. C. (1998). "An Application of the Bootstrap Method to the Simultaneous Equations Model of the Demand and Supply of Audit Services." *Contemporary Accounting Research* 15:83–99.

Estrella, A. (1998). "A New Measure of Fit for Equations with Dichotomous Dependent Variables." *Journal of Business and Economic Statistics* 16:198–205.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: John Wiley & Sons.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. 2nd ed. London: Chapman & Hall.

Genz, A. (1992). "Numerical Computation of Multivariate Normal Probabilities." *Journal of Computational and Graphical Statistics* 1:141–150.

Geweke, J. (1995). *Monte Carlo Simulation and Numerical Integration*. Staff Research Report 192, Federal Reserve Bank of Minneapolis.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*. Cambridge: Cambridge University Press.

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987). "Generalized Residuals." *Journal of Econometrics* 34:5–32.

Greene, W. H. (1997). *Econometric Analysis*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Greene, W. H. (2001). "Fixed and Random Effects in Nonlinear Models." Department of Economics, Leonard N. Stern School of Business, New York University.

Gregory, A. W., and Veall, M. R. (1985). "On Formulating Wald Tests for Nonlinear Restrictions." *Econometrica* 53:1465–1468.

Hajivassiliou, V. A. (1993). "Simulation Estimation Methods for Limited Dependent Variable Models." In *Econometrics*, edited by G. S. Maddala, C. R. Rao, and H. D. Vinod. Vol. 11 of Handbook of Statistics, 519–543. New York: Elsevier Science.

Hajivassiliou, V. A., and McFadden, D. L. (1998). "The Method of Simulated Scores for the Estimation of LDV Models." *Econometrica* 66:863–896.

Hajivassiliou, V. A., McFadden, D. L., and Ruud, P. A. (1996). "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results." *Journal of Econometrics* 72:85–134.

Halton, J. H. (1960). "On the Efficiency of Certain Quasi-random Sequences of Points in Evaluating Multi-dimensional Integrals." *Numerische Mathematik* 2:84–90.

Heckman, J. J. (1978). "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46:931–959.

Hinkley, D. V. (1975). "On Power Transformations to Symmetry." *Biometrika* 62:101–111.

Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. (1982). "On the Estimation of Technical Efficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics* 19:233–238.

Kim, M., and Hill, R. C. (1993). "The Box-Cox Transformation-of-Variables in Regression." *Empirical Economics* 18:307–319.

Kumbhakar, S. C., and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. New York: Cambridge University Press.

Lee, L.-F. (1981). "Simultaneous Equations Models with Discrete and Censored Dependent Variables." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden, 346–364. Cambridge, MA: MIT Press.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–142. New York: Academic Press.

McFadden, D. (1981). "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden, 2–50. Cambridge, MA: MIT Press.

McKelvey, R. D., and Zavoina, W. (1975). "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4:103–120.

Meeusen, W., and van den Broeck, J. (1977). "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18:435–444.

Morokoff, W. J., and Caflisch, R. E. (1995). "Quasi-Monte Carlo Integration." *Journal of Computational Physics* 122:218–230.

Mroz, T. A. (1987). "The Sensitivity of an Empirical Model of Married Women's Work to Economic and Statistical Assumptions." *Econometrica* 55:765–799.

Mroz, T. A. (1999). "Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome." *Journal of Econometrics* 92:233–274.

Nawata, K. (1994). "Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator." *Economics Letters* 45:33–40.

Parks, R. W. (1967). "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated." *Journal of the American Statistical Association* 62:500–509.

Phillips, C. B., and Park, J. Y. (1988). "On Formulating Wald Tests of Nonlinear Restrictions." *Econometrica* 56:1065–1083.

Powers, D. A., and Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.

Rivers, D., and Vuong, Q. H. (1988). "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39:347–366.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms." *Annals of Applied Probability* 7:110–120.

Roberts, G. O., and Rosenthal, J. S. (2001). "Optimal Scaling for Various Metropolis-Hastings Algorithms." *Statistical Science* 16:351–367.

Schervish, M. J. (1995). *Theory of Statistics*. New York: Springer-Verlag.

Sloan, I. H., and Woźniakowski, H. (1998). "When Are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals?" *Journal of Complexity* 14:1–33.

Spanier, J., and Maize, E. (1991). "Quasi-random Methods for Estimating Integrals Using Relatively Small Samples." *SIAM Review* 36:18–44.

Stacy, E. W. (1962). "A Generalization of the Gamma Distribution." *Annals of Mathematical Statistics* 33:1187–1192.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge: Cambridge University Press.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Subject Index

Syntax Index