

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

This chapter provides step by step instructions on how to define and estimate each of the three types of LC models (Cluster, DFactor or Regression) and also details the various model options. You need not follow all 10 steps in order to perform an analysis. For example, if you choose to maintain the default Technical and Output option settings, then you would skip Steps 8 and 9.

These steps are illustrated in Tutorial #1 (Chapter 7) for simple LC Cluster models.

Step 1: Load Your Data into Latent GOLD



▷ From the menus choose: File → Open

Or select the Open icon  from the toolbar.

▷ In the Open dialog box, select the type and name of the file you want to open.

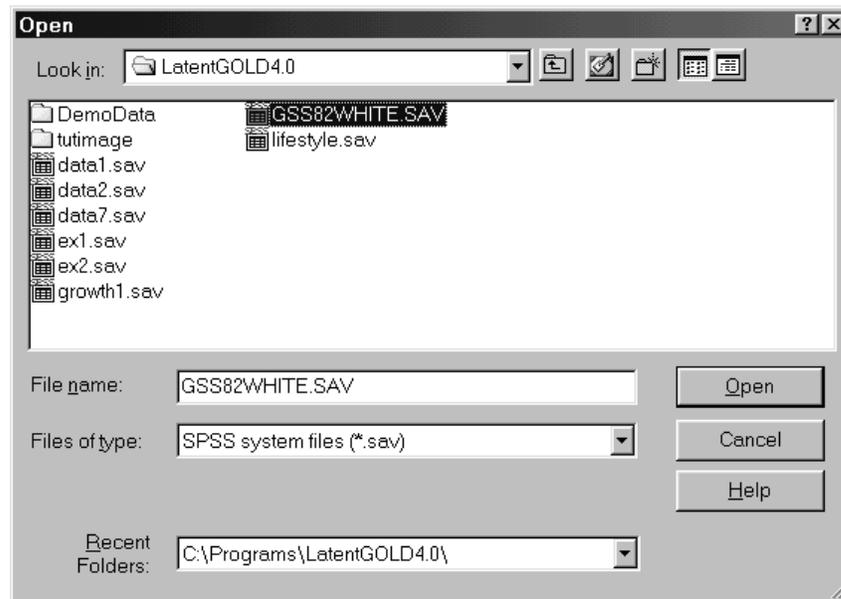


Figure 5-1. File Open Dialog Box

▷ Click Open.



For more detailed information on File Open options and on the types of files Latent GOLD can read, see Chapter 3: Data Files and Formats.

Step 2: Select the Type of Model

If you open a data file, the Outline pane will contain the data file name plus the default model name ('Model1'). If you open a model definition (lgf) file, the settings for all models previously saved in this definition file will be loaded into the program and the Outline pane will contain a list of model names for each of these models.

Assuming that you have opened a data file and no models have yet been estimated for the correct data file, the next step is to use the Model Menu to choose from one of the three analysis modules:

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

| MODEL TYPE | ANALYSIS PERFORMED |
|------------|--|
| Cluster | LC cluster analysis |
| Dfactor | Dfactor analysis |
| Regression | LC segmentation/regression analysis |
| Choice | LC choice analysis (requires Latent GOLD Choice license) |



A quick way to open the Model Menu is to right click on a model name. If no models have yet been estimated for the current data file, the default name for the new model to be estimated (Model1) will be the only model name that appears in the Outline Pane.

▷ Right click on the model name and the Model Menu appears:

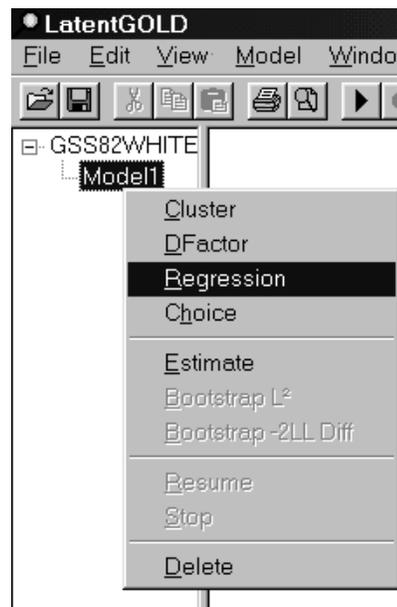


Figure 5-2. Model Menu

The Model Menu options are organized into 4 sections. Those used in step 2 are listed in the top section. The other options are discussed in step 10.

The options for Model Type:

Model

Cluster/DFactor/Regression

▷ Choose from one of these to select the type of model to be estimated.

Once you select a model type, the Variables Tab of the appropriate Analysis Dialog Box opens. For example, Figure 5-3 shows the Variables Tab associated with the Cluster model type.

As an alternative to beginning your model setup from scratch, you may begin by altering the settings of a previously defined model. If one or more models have already been estimated for a data file, or if you opened a previously saved .lgf files you can right click on its name to display the Model Menu, which will contain a checkmark next to the type of model currently associated with that name. You may select that model type again or select one of the other model types. To maintain the model type, instead of a right click, you may simply double click on the name of the model and the Variables Tab corresponding to the associated model type opens with the current settings for that model.

If you right click on the name of a new model created automatically at the bottom of the list of model names, the checkmark appears next to the last model type estimated on the associated data file.

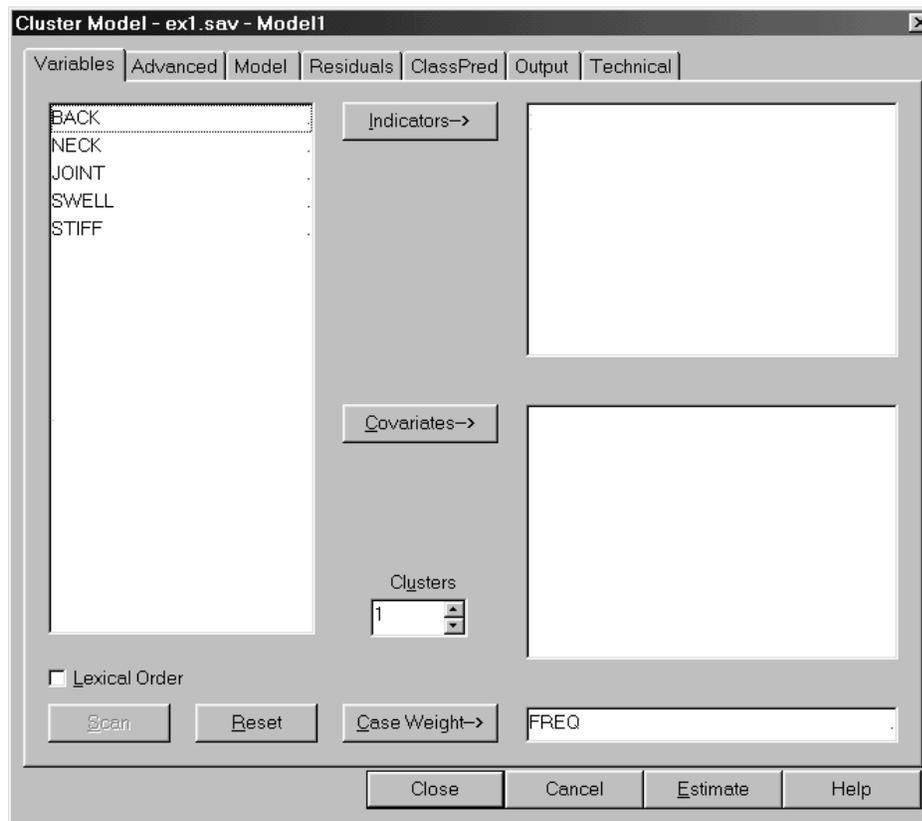


Figure 5-3: Variables Tab for an LC Cluster Model

The next several steps show how to specify the desired model using the various tabs of the Analysis Dialog Box. Specifically, use of the setup options in the Variables Tab are described in Steps 3, 4, 5 and 6, the Models, Residuals, and (optional) Advanced Tab in Step 7, the Technical Tab in Step 8 and the Output Tab in Step 9.

The appearance of the various tabs differs by Model Type. However, regardless of the type of model, the Title bar (above the main menu in any open Analysis dialog box) lists the model type, data file name and model name as identifying information. In addition, the following buttons appear below any Analysis Dialog Box.

Close. Closes the analysis dialog box but retains any changes you have made since opening the dialog box.

Cancel. Closes the analysis dialog box without retaining any changes you have made since opening the dialog box.

Estimate. Initiates estimation of the model based on the current model option settings

Help. Provides help associated with the options for the particular model type.

Step 3: Select Variables for the Analysis (Variables Tab)

The appearance of the Variables Tab differs depending on the type of model you have selected. The primary differences are:

- **For LC Cluster and DFactor models, one or more variables must be selected from the variables list box to be used as Indicators, also known as Dependent Variables. For LC Regression models, one and only one Dependent Variable is used and one or more variables may be selected as Predictors.**
- **The Regression Module differs from the Cluster and DFactor Modules in that it can accommodate multiple records (repeated measures) for one or more cases. Thus, the use of a case ID and a Replication Weight is permissible in Regression.**

RESET button

For any model type, at any time during the model setup process, the Reset button can be used to restore the default settings (maintaining only the specified number of classes). Upon selecting Reset, all selected variables (including the Case Weight variable) are returned to their original position in the variables list box. The Reset Button is located in the lower left portion of the Variables Tab.

In addition, for any model type, the Variables Tab can be used to specify a variable as a Case Weight, and one or more variables as Covariates. The specification of variables as a Case Weight or Covariate(s) is optional.

SPECIFYING A CASE WEIGHT (OPTIONAL)

You may assign one variable to be used as a case weight (usually a frequency variable) in either a Cluster, DFactor or Regression model. See Tutorial #1 in Chapter 7 for an example that describes the use of a case weight.

For an SPSS .sav file in which a variable is designated as a weight variable (according to the SPSS data dictionary), this variable appears in the Case Weight box automatically upon opening the .sav file.



To specify a case weight:

- ▷ In the Variables Box, highlight the variable you want as your case weight.
- ▷ Click "Case Weight" to move this variable to the Case Weight Box.

COVARIATES (OPTIONAL)

Covariates are variables that may be used to describe or predict (rather than to define or measure) the latent classes and if Active, to reduce classification error. For example, they are often used to profile the latent classes in terms of demographic or other exogenous variables. Covariates may be treated as Nominal or Numeric and may be Active or Inactive (see Step 5: Set Scale Types). Select any variables you want to use as covariates. For a formal distinction between covariates (zcov), predictors (zpred), indicators and dependent variables (y), see section 2.1 of the Technical Guide.



Group level covariates (GCovariates) If 2 or more group latent classes (GClasses) have been included in the model, any Active covariates can be specified as group level covariates (GCovariates) for describing these GClasses. For details, see the Advanced Modeling Options in Step 7.

INDICATORS (CLUSTER AND DFACTOR MODELS)

Indicators are dependent variables that are used to define or measure the latent classes in a LC Cluster model, or latent variable(s) in a DFactor model. Indicators may be treated as Nominal, Ordinal, Continuous, Poisson Count, or Binomial Count (see Step 5: Set Scale Types).

Select one or more variables from the variables list box to be used as **Indicators** (required).

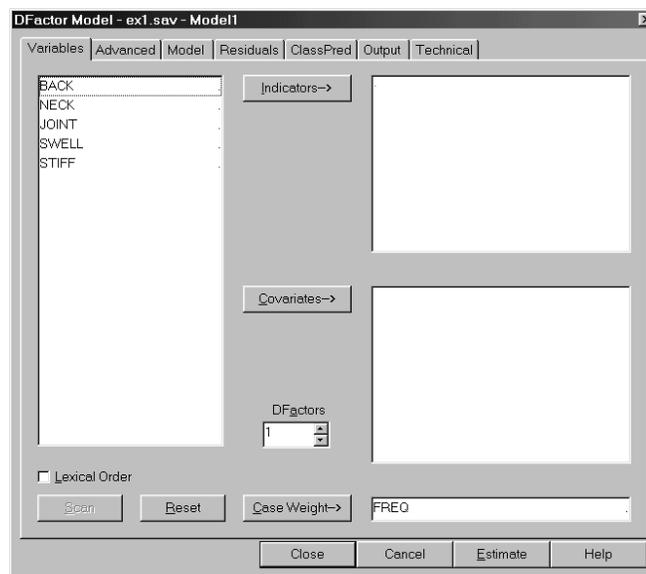


Figure 5-4: Variables Tab for a DFactor Model

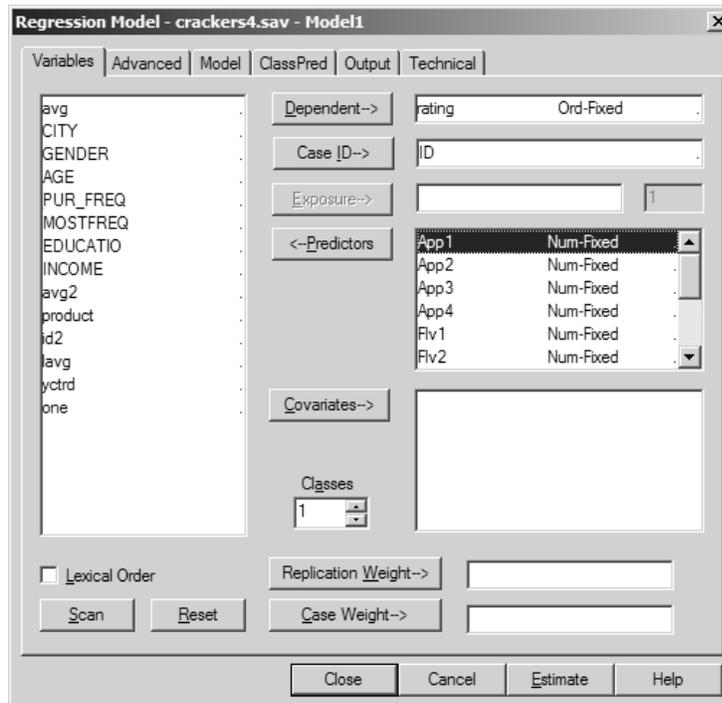


Figure 5-5: Variables Tab for LC Regression Model

DEPENDENT (REGRESSION MODELS)

Assign one variable to be used as the dependent variable. Dependent variables may be specified as Nominal, Ordinal, Continuous, Count or Binomial Count (see Step 5: Set Scale Types). The Continuous scale type causes the normal distribution to be used resulting in a **linear** LC Regression model.

The appropriate model is estimated according to the dependent variable scale type:

Continuous. Linear regression (with normally distributed residuals)

Dichotomous (specified as nominal, ordinal, or a binomial count). Binary logistic regression

Nominal (with more than 2 levels). Multinomial logistic regression

Ordinal (with more than 2 ordered levels). Adjacent-category ordinal logistic regression

Count. Log-linear Poisson regression

Binomial Count. Binomial logistic regression model

The subtype **censored continuous** yields a tobit regression model. The subtypes **truncated continuous**, **truncated count**, and **truncated binomial** count yield truncated versions of the linear, log-linear Poisson, and binomial logistic regression model, respectively.

PREDICTORS (REGRESSION MODELS, OPTIONAL)

Select any variable(s) to be used as predictors of the dependent variable. Predictors may be treated as Nominal or Numeric (see Step 5: Set Scale Types) and as Class Independent or Class Dependent.

If no predictors are selected, the model will contain an intercept only.

EXPOSURE (REGRESSION MODELS FOR COUNTS, OPTIONAL)

The Exposure button is active only if the scale type for the dependent variable has been specified to be Count or Binomial Count. (For other scale types, no exposure variable is used.)

For dependent variables specified as Count or Binomial Count, the exposure is specified by designating a variable as the exposure variable or, if no such variable is designated, by entering a value in the exposure constant box which appears to the right of the Exposure variable box. The use of an exposure variable allows the exposure to vary over cases.

By default, the value in the Exposure constant box is 1, a value often used to represent the Poisson exposure. To change the exposure constant, highlight the value in the exposure constant box and type in the desired value. If an exposure variable is selected, this option is not available.

When the scale type is specified as Binomial Count, the value of the dependent variable represents the number of 'successes' in N trials. In this case, the exposure represents the number of trials (the values for N), and hence should never take on a value lower than the value of the dependent variable and hence typically should be higher than the default constant of 1. Before the actual model estimation, Latent GOLD checks each case and will provide a warning message if this condition is not met for one or more cases. An exposure variable should be designated if the number of trials is not the same for all cases.

Regression Models with Repeated Measurements

For LC Regression models where more than one record is included for at least one case, a Case ID variable must be specified and a Replication Weight may be used.

CASE ID VARIABLE (REGRESSION MODELS, OPTIONAL)

If the data file contains only 1 record per case, no Case ID variable is required. For data files in which one or more cases have multiple records (e.g. repeated measures), a variable must be assigned as a Case ID variable to uniquely identify each case. For a regression example with repeated measures is given in Tutorial #3 (see Chapter 7).

To assign a variable as a Case ID, select that variable from the Variable List Box and click the Case ID button. The variable is moved to the Case ID box..

For data files containing multiple records per case:

- A case weight (if used) should take on the same value for each record associated with the same Case ID.
- Any covariates should take on the same value for each record associated with the same Case ID.
- If these conditions are not met, a warning message is produced during the file Scan (Step 4).

REPLICATION WEIGHT (REPEATED MEASURES REGRESSION MODELS, OPTIONAL)

Once a variable is selected as a Case ID, the Replication Weight button becomes active. To assign a variable as a Replication Weight, select that variable from the Variable List Box and click the Replication Weight button. The variable is moved to the Replication Weight box. A common application of replication weights is in the estimation of certain kinds of allocation models, where respondents assign a fixed number of points to each of J alternatives. For each case, the assigned points are used as replication weights to weight each of J responses. A weighted multinomial logit model is estimated. See Section 5.1 of Technical Guide for further details about replication weights.

Step 4: Scan the Data File (Variables Tab)

The next step is to ensure the integrity of the data by checking the number of cases, number of records, the observed distribution of each variable, any category labels that may be included on the file, and that certain consistency checks are satisfied. This is accomplished with the File Scan, which also issues warning messages if certain inconsistencies are found.

- ▷ Once the desired variables are selected, click Scan.

This option scans the data file and identifies each distinct category or value for all variables included in a model. Prior to a data file being scanned, only variable names will be listed in the target list box. After scanning, a message appears in the status bar indicating the number of records in the file, and the number of distinct categories (or values) appears to the right of the variable names in the target list boxes. If a Scan has not been selected before clicking Estimate, a scan will automatically be performed by Latent GOLD prior to beginning the estimation algorithm, and default settings will be used to scale the variables in the model.

Following a Scan, a "Replication Error" warning message appears if a covariate or case weight is not constant for all records corresponding to the same Case ID. The option to "Estimate anyway?" is given at the end of the error message. If Yes is selected, Latent GOLD uses the value on the first record as the desired constant.

VIEWING CATEGORY LABELS, FREQUENCY COUNTS AND SCORES FOR A VARIABLE

After scanning, double click on a variable name in any of the target list boxes to open the Variables dialog box where category labels, frequency counts and any scores assigned to the variable are displayed. Any scores assigned to Nominal variables are used for descriptive purposes only. Scores associated with ordinal indicators may be changed. For further information about score options for Ordinal variables, see How to Specify Category Scores for an Ordinal Variable.

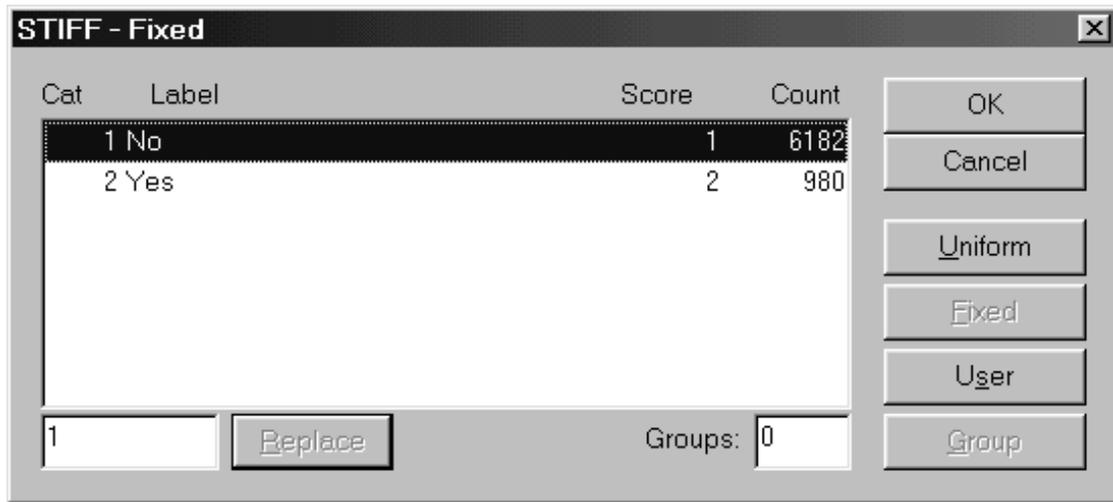


Figure 5-6. Category labels for the variable STIFF

MISSING VALUES

Latent GOLD recognizes observations as missing on a particular variable if they were assigned a missing value code. In SPSS (for .sav formatted files) missing values may be specified as system missing or user missing. In text formatted files a missing value code is represented by '.'

During the Scan process, all records pertaining to such cases will be deleted. That is, by default, missing values are eliminated using list-wise deletion. After the scan, a message in the status bar indicates the number of records deleted because of missing values.

By default, cases containing missing values on any indicator, dependent variable, predictor or active covariate are excluded from the analysis. The treatment of missing values may be changed to allow cases containing missing values to be included in the analysis for the various options. To include such missing values in an analysis, see Missing Values in Step 8.

REDUCING THE NUMBER OF CATEGORIES FOR A VARIABLE

After verifying the integrity of your data, you may wish to reduce the size of the data file to speed up estimation, especially if you are estimating a model with many parameters. The Group option can be used to reduce the number of values taken on by a variable by combining adjacent categories (in most cases, you would not want to group a Nominal variable in this manner). The score assigned to a group of adjacent categories is the average of the original (default) scores assigned to the original categories. The grouping algorithm groups adjacent categories that have fewer observations before grouping categories having larger number of observations towards the goal of having approximately equal sized groups.

The grouping used in Latent GOLD is identical to the RANK command used in SPSS with the number of groups equal to the NTILES subcommand.



How To Group Variable Categories

- ▷ Following a file Scan, double click on a variable name to open the Variables dialog box.
- ▷ Type the number of desired groups into the Groups list box.
- ▷ Click Group.

The variables dialog box changes to show the new groups. The actual number of groups will be less than or equal to the number specified. After grouping, the label field contains the range for the original categories that have been grouped together, and the scores field contains the average of the scores originally assigned to cases in that group. Specifying 0 or a number equal to the number of original categories restores the variable to its original number of categories.

| Cat | Label | Score | Count |
|-----|---------|-----------|-------|
| 1 | 1 - 8 | 30.866667 | 15 |
| 2 | 9 - 15 | 45.833333 | 12 |
| 3 | 16 - 22 | 54.588235 | 17 |
| 4 | 23 - 25 | 58.833333 | 12 |
| 5 | 26 - 30 | 62.785714 | 14 |
| 6 | 31 - 36 | 68.285714 | 14 |

30.866667 Groups: 6

Figure 5-7. Grouping the variable AGE

The 36-category AGE variable has been merged into 6 groups. New group #1 consists of the original AGE categories 1-8 (abbreviated as "1 1-8"), the 8 youngest ages. The score for this group (displayed in the score field) is 30.87, computed as the average of the original AGE scores assigned to the 15 cases in this group.



To accept the category grouping and new scores

- ▷ Click OK

The new number of categories will now be listed next to the variable name in the target list along with the letter 'G' (signifying the variable has been grouped).

Note: For nominal and ordinal variables, the labels for the grouped categories consist of indices of the original categories. For continuous variables and counts, the label corresponds to the range of values.



To Restore a Grouped Variable to its Default Setting

- ▷ Highlight the grouped variable name in the target list.
- ▷ Double click the variable name to open the Variables dialog box.
- ▷ Type '0' or the number of original categories in the Groups list box.
- ▷ Click OK.

WARNING!

Use of the Group option will cause any previously set user scores to be ignored. User scores can be assigned to the new categories of a grouped variable to replace the average scores.

Step 5: Set Scale Types (Variables Tab)

Scale Types determine the distributional assumptions and structural form of the model to be estimated. There are several different scale types (Nominal, Ordinal, Continuous, Count, Binomial Count, Numeric) and subtypes (Truncated, Censored, Standard, Zero-Inflated and Overdispersed) that may be assigned to a variable depending upon the type of variable (indicator, covariate, dependent, predictor). The default for indicators and dependent variables is Ordinal when the variable contains numeric (quantitative) values, and Nominal when the variable contains character codes (string variables). The default for covariates and predictors is Numeric when the variable contains numeric (quantitative) values, and Nominal when the variable contains character codes (string variables).



To assign or change a scale type:

- ▷ Highlight the variable name(s) in a target list.
- ▷ Right click to open a menu with the selections for that variable type.
- ▷ Select a scale type.

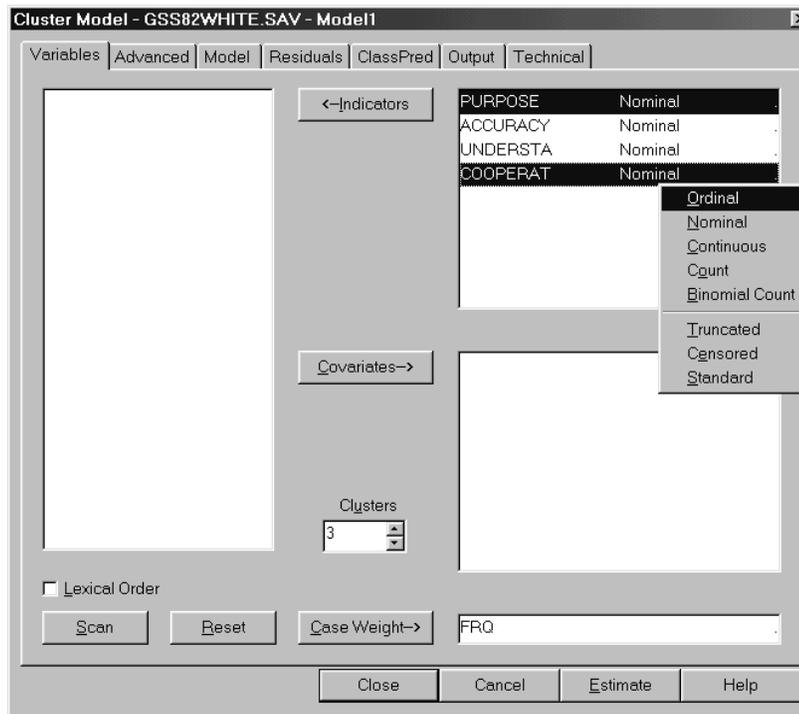


Figure 5-8. Selecting Scale Types

The selected scale type will appear to the right of the variable name. If Ordinal is selected, by default, the scores will be treated as Fixed at the values on the data file and the label "Ord-Fixed" appears to change one or more of the category scores to different values, see How to Specify Category Scores for an Ordinal Variable.

Scale Types

Each scale type is listed below along with related details.

ORDINAL (INDICATORS, DEPENDENT)

This setting should be used for categorical variables where the categories are ordered (either from high to low or low to high). If the dependent variable in an LC Regression model is set to Ordinal, the adjacent category logit model, also known as the baseline category logit model (see e.g. Magidson, 1996, 1998), is specified. For an Ordinal variable, the scale type will appear as 'Ord-Fixed', 'Ord-Uniform' or 'Ord-User' depending upon which scores are used (fixed, uniform or user scores):

- 1) **Fixed (default).** Either the original values contained in the data file or array, or, if no scores were specified, uniform scores are assigned to the categories.
- 2) **Uniform.** The variable is assigned fixed scores that are equidistant and have a low score of 0 and a high score of 1. For example, for a 3 category variable, the uniform scores assigned are 0, 0.5, and 1.
- 3) **User Scores.** The user specifies category scores.

How to Specify Category Scores for an Ordinal Variable

- ▷ After scanning your data file, double click on the variable name to open the Variables dialog box.



To specify uniform scores,

- ▷ Click Uniform.
- ▷ Click OK to accept the new scores and close the Variables dialog box.
- ▷ The label for the variable in the Variables Tab changes to "Ord-Uniform"



To specify user scores,

- ▷ Double click on any score you wish to change.
- ▷ Type in a value for the new score.
- ▷ Click Replace.
- ▷ Click OK to accept the new scores and close the Variables dialog box.
- ▷ The label for the variable in the Variables Tab changes to "Ord-User"

If the user inputs scores, and then later resets them by clicking Uniform or Fixed, clicking User will reset scores to their most recent user setting.



To restore the default setting:

- ▷ Click Fixed.
- ▷ Click OK to accept the new scores and close the Variables dialog box.
- ▷ The label for the variable in the Variables Tab is reset to "Ord-Fixed"

NOMINAL (INDICATORS, COVARIATES, DEPENDENT, PREDICTORS)

This setting should be used for categorical variables where the categories have no natural ordering. When estimating an LC Regression model, if the dependent variable is set to Nominal, the multinomial logit model is used.

NUMERIC (COVARIATES, PREDICTORS)

This setting should be used for an ordinal or continuous covariate or predictor.

ACTIVE (COVARIATES)

Specifying a covariate as active influences the definition of the latent classes.

INACTIVE (COVARIATES)

Specifying the covariate as **inactive** guarantees that its inclusion in the model has no influence on the model parameter estimates: An inactive covariate is not part of the specified model.



ADVANCED: GROUP (COVARIATES)

If a multilevel model is specified, group level covariates (GCovariates) may also be included in the covariate box. To distinguish covariates at the group level from ordinary covariates, select the Group setting. A <G> appears next to the group level covariate. All group level covariates are **active**.

CONTINUOUS (INDICATORS, DEPENDENT)

This setting should be used when the variable is continuous. When estimating an LC Regression model, if the dependent variable is set to Continuous, the normal linear Regression model is used.

COUNT (INDICATORS, DEPENDENT)

This setting should be used when the variable represents Poisson counts. When estimating a LC Regression model, if the dependent variable is set to Count, the Poisson model is used and you can also specify an additional variable to be used as an exposure (see Exposure)

When estimating a LC Cluster or DFactor model, the exposure is set to 1 for all indicators specified as having the Count scale type.

BINOMIAL COUNT (INDICATORS, DEPENDENT)

This setting should be used when the variable represents binomial counts. When estimating a LC Regression Model, if the dependent variable is set to Binomial Count, the binomial model is used and you can also specify a variable to be used as an exposure (see Exposure). During the scan, the program checks to make sure that the exposure, if specified, is larger than any observed count.

When estimating a LC Cluster or DFactor model, for any indicator(s) specified as having the Binomial Count scale type, the exposure is computed automatically by the program to be the largest count observed for the indicator(s).

Note: The exposure can easily be set to a value larger than the largest observed count by means of the following trick: Add one case with xvalues equal to the desired exposure values for all indicators. Include a very small case weight (say 1.0E-50) to the data file for this additional case.

Scale Subtypes

Four scale subtypes are listed below along with related details.

STANDARD (INDICATORS, DEPENDENT)

By default, this setting is used for indicators and dependent variables. This setting means that the standard scale type is used for this variable as opposed to one of the special subtypes described below.

TRUNCATED (COUNT, BINOMIAL COUNT, CONTINUOUS)

Truncated indicates that only cases with a value larger than 0 are in the sample. It can be applied with an Indicator/Dependent that is a Poisson or Binomial Count, or a Continuous normally distributed variable. The likelihood function is then constructed using truncated variants of the Poisson, binomial, and normal distribution.

CENSORED (CONTINUOUS)

Censored means that all cases with a "true" value smaller than 0 have the same observed value of 0; that is, the variable is assumed to be left censored at 0. Censored can be used with a Continuous Indicator/Dependent, in which case we use assume a left censored normal distribution for the variable concerned.

ZERO INFLATED (REGRESSION ONLY)

Zero Inflated (ZIN) means that one additional latent class with a mean equal to zero is automatically added to the specified model. For Regression models with the dependent variable being a count, these specifications yield the well known Zero Inflated variants of the Poisson and binomial count regression models. In the case of a Continuous dependent variable, it yields a Zero Inflated variant for left censored normal regression, sometimes referred to as censored-inflated regression. In other words, for continuous dependent variables, Zero Inflated also implies Censored.

With Nominal and Ordinal dependent variables, (as well as with Choice, Rating and Ranking specification in Latent GOLD Choice), one ZIN class is added for each category of the Dependent. Each ZIN class responds with probability 1 into a particular category (and with probability 0 into the other categories). These classes are sometimes referred to as stayer classes (in a mover-stayer model) or brand-loyal classes (in a brand-switching model).

OVERDISPERSED (COUNT AND BINOMIAL COUNT IN REGRESSION)

Overdispersion is a common phenomenon in count data. It means that, as a result of unobserved heterogeneity, the variance of the count variable is larger than estimated by the Poisson (binomial) model. The overdispersed option makes it possible to account for unobserved heterogeneity by assuming that the rates (success probabilities) follow a gamma (beta) distribution. This yields a negative-binomial model for overdispersed Poisson counts and a negative-binomial model for overdispersed binomial counts. Note that this option is conceptually similar to including a normally distributed random intercept in a regression model for a count variable.

The overdispersion option is useful if one wishes to analyze count data using mixture or zero-inflated variants of (truncated) negative-binomial or beta-binomial models (Agresti, 2000; Long, 1997; Simonoff, 2003). The negative-binomial model is a Poisson model with an extra error term coming from a gamma distribution. The beta-binomial model is a variant of the binomial count model that assumes that the success probabilities come from a beta distribution. These models are common in fields such as criminology, political sciences, medicine, biology, and marketing.

Step 6: Specify the Number of Latent Classes (Variables Tab)

Use the Variables Tab to specify the number of classes for your model. In the Cluster Module, a class is called a Cluster. In the DFactor Module, the ordered latent classes associated with each DFactor are called 'levels'. A model may contain several DFactors, each of which may have a different number of levels.

SPECIFYING THE NUMBER OF CLUSTERS FOR LC CLUSTER MODELS

The box labeled Clusters is located beneath the Indicators button in the Variables Tab.

INDICATORS (CLUSTER AND DFACTOR MODELS)

Indicators are dependent variables that are used to define or measure the latent classes in a LC Cluster model, or latent variable(s) in a DFactor model. Indicators may be treated as Nominal, Ordinal, Continuous, Poisson Count, or Binomial Count (see Step 5: Set Scale Types).

Select one or more variables from the variables list box to be used as Indicators (required).

- ▷ Enter a number greater than 0. You may also specify a range to estimate several models. For example, enter "1-4" to estimate four different latent class models containing 1, 2, 3, and 4 clusters respectively.

SPECIFYING THE NUMBER OF DFATORS FOR DFACTOR MODELS

The box labeled DFactors is located below the Covariates button in Figure 5-4. Enter a number greater than 0. (The range option is not available when specifying the number of DFactors.)

By default, each DFactor consists of 2 levels. To increase the number of levels for one or more DFactors, use the Model Tab and set the number of levels desired for each factor (See Step 7: Set Restrictions and Other Model Options). The maximum number of levels that can be specified for a DFactor is 20.

SPECIFYING THE NUMBER OF CLASSES FOR LC REGRESSION MODELS

The box labeled Classes is located beneath the Covariates button in Figure 5-5. Enter a number greater than 0. Separate regression models will be estimated for each segment. If a range is specified for the number of classes, such as '1-4', separate sets of regression models will be estimated, the first representing a 1-class regression model (the traditional regression which assumes a single homogeneous population). For the other models, separate regressions are estimated for 2, 3 and 4 classes (segments) respectively.



ADVANCED. SPECIFY THE NUMBER OF CFATORS, GCLASSES, GCFACTORS

An advanced option allows for the inclusion of up to 3 continuous factors (CFactors) in a model. In addition, a multilevel extension of a model may be estimated which involves specifying a group-level ID variable. In the case of multilevel models, 2 or more group-level latent classes (GClasses) and up to 3 group-level CFactors (GCFactors) may be specified. For further details, see Step 7: Advanced Tab.

Step 7: Set Restrictions and Other Model Options

Following Step 6, you may choose to estimate your model (Step 10), view the results and then impose (post-hoc) restrictions on the parameters to achieve a more parsimonious and interpretable model. Alternatively, you may choose to impose a priori restrictions and/or relax certain restrictions that are imposed by default. The most common restrictions are:

- zero restrictions:** restrict to zero any parameter estimate that is not statistically significant
- class independence restrictions:** restrict certain parameter estimates to be identical for each latent class
- offset:** specify certain parameters to be equal to one (Regression)
- equal effects:** equate parameters across indicators (Cluster and DFactor)

The additional effects that can be included are:

- associations between DFactors (DFactor model)
- CFactor1 affecting the predictor coefficients (Regression Advanced),
- GClasses and GCFactor effects on the indicators (Cluster/DFactor Advanced)
- GClasses and GCFactors affecting the intercept and the predictor coefficients (Regression)

Advanced),

- GClasses and GCFactors affecting the covariate effects in the regression model for the Clusters, Classes, or DFactors (Advanced)

These are a few of the kinds of restrictions that may be applied using the Model Tab. Some additional kinds of restrictions and other advanced options are available using other tabs:

ClassPred Tab: may be used to assign with certainty one or more designated cases to belong to a particular latent class or latent classes. This Known Class option amounts to indicating to which classes cases or subsets of cases may not belong.

Residuals Tab (Cluster and DFactor models only): may be used to include direct relationships between indicators (associations/covariances) and direct effects of covariates on indicators in a model.



Advanced Tab (available in the Advanced version of Latent GOLD): may be used to

- add group- level continuous latent variables (GCFactors) and/or a group level nominal latent variable (GClasses) in a multilevel extension of a model, as well as indicate which coefficients in the logistic regression model for the Clusters, DFactors, and Classes are random effects (differ across GClasses or are affected by GCFactors)
- add continuous factors (CFactors) to specify factor analytic, item response theory (IRT), or random-effects regression models for 2-level data
- set options to incorporate the sampling design if it deviates from simple random sampling.

This section details the use of each of these options.

Model Tab: Applying (or Relaxing) Parameter Restrictions

The appearance of the Model Tab and the available options differ depending upon the type of model selected.

LC CLUSTER MODEL

The LC Cluster Model Tab allows you to change the number of clusters, and to restrict the following parameters in various ways:

- Beta effects of the latent variable on the selected indicators
- Gamma effects of selected (active) covariates on the latent variable
- Error variance parameters (for continuous variables only)
- Error covariance parameters (for continuous variables only)
- Advanced: Lambda effects of selected CFactors and GCFactors on selected indicators; beta effects of the GClasses on selected indicators; gamma effects of selected group covariates on the GClasses

The Name List Box appears in the left-most portion of the Model Tab. It contains the name 'Clusters' at the top, followed by the specific cluster names ('Cluster1', 'Cluster2', ...) for each cluster specified under Step 6. If the range option was used to specify a range of cluster models to be estimated, no specific cluster names will be listed in this box.



At the bottom of this list will be specific names for each CFactor specified ('CFactor1', 'CFactor2', 'CFactor3') if any CFactor effects have been specified in the Advanced Tab. For multilevel models, the list also contains GClasses and/or GCFactors.

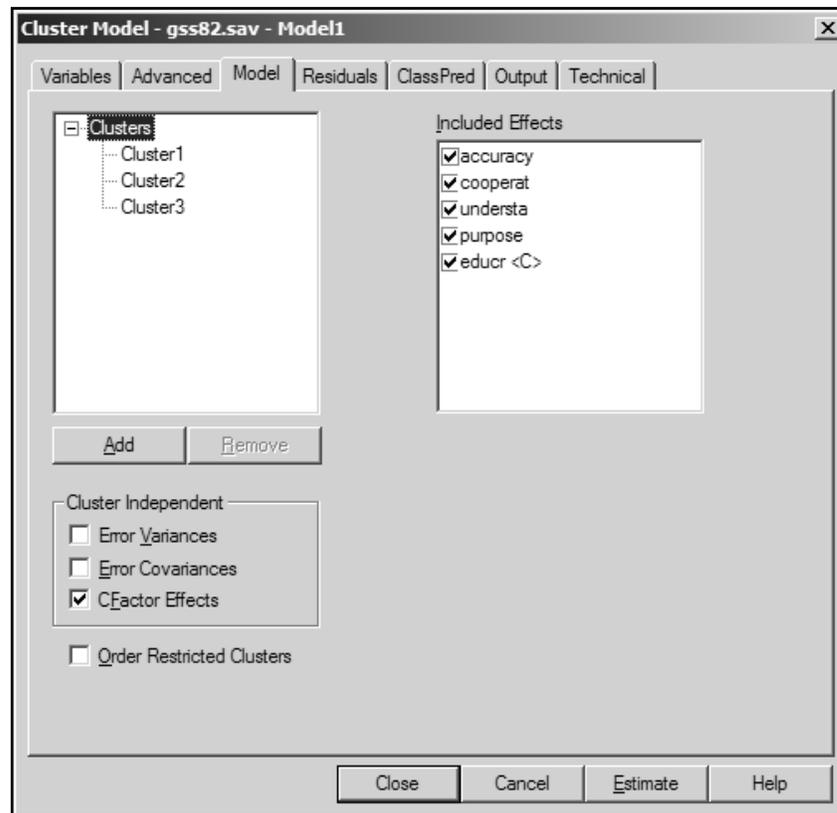


Figure 5-9. Model Tab for LC Cluster Model

CHANGING THE NUMBER OF CLUSTERS

The Add and Remove buttons can be used as an alternative to the clusters box in the Variables Tab to set the number of clusters. To change the number of clusters (from that specified in Step 6), you can increase or decrease the number of clusters by selecting 'Clusters' (or one of the cluster names) and clicking Add to increase the number of clusters. You can also select one of the cluster names and click the Remove button to decrease the number of clusters.

Note: The Add and Remove buttons will not be active if the Range option was used in Step 6 to specify a range of clusters.

IMPOSING ZERO RESTRICTIONS

The Included Effects box contains all indicators and active covariates that are included in the current model. Variable names for the active covariates are followed by the characters '<C>'. By default, each of these variable names is preceded by a check (check box equals on) to indicate that a set of beta effects will be estimated for the associated indicator or that a set of gamma effects will be estimated for the associated covariate. To restrict any of these sets of effects to zero, click in the check box and the check is removed (check box equals off).



If 1 or more CFactors is included in a model, names for these CFactors ('CFactor1', 'CFactor2', 'CFactor3') appear beneath the cluster names. By default, for each CFactor, lambda effects will be estimated for all indicators included in the model. For a selected CFactor, click on the name of that CFactor and the Included Effects box shows the lambda effects to be estimated for that CFactor. To restrict any of these lambda effect(s) to zero, click in the check-box of the indicator(s) and the check is removed.

Note: Since it is not possible to specify regression models for CFactors, the check-box for any Covariates is inactive for CFactor effects.



If GClasses or one or more CFactors is included in a model, names for these appear beneath the cluster names. By default, for GClasses and each GCFactor, beta(g) and lambda(g) effects will be excluded for all indicators included. For GClasses or a selected CFactor, click on its name and the Included Effects box shows the effects to be estimated for that latent variable. To add any of these beta/lambda effect(s), click in the check-box of the indicator(s) and the check is added.

Note: Since it is not possible to specify regression models for GCFactors, the check-box for any Covariates is inactive for GCFactor effects. Since Group Covariates can be used as predictors in the regression model for the GClasses, the check boxes for the corresponding gamma(g) effects are active and by default on (effect included).

For cluster effects:

Make sure that the name 'Clusters' (or one of the cluster names) is highlighted in the left-most box

- Removing a checkmark for selected indicators, restricts a set of beta effects to zero.
- Removing a checkmark for selected active covariates, restricts a set of gamma effects to zero.

Note: The set of beta (gamma) effects restricted to zero are associated with all clusters. Zero restrictions can not be used to restrict betas or gammas to zero for certain classes while allowing them to be estimated for other clusters.



For CFactor effects:

Make sure that the name of the CFactor (e.g., 'CFactor1') is highlighted in the left-most box

- Removing a checkmark for selected indicators restricts lambda effects to zero.

- The check-box is inactive for covariates.
- By default, effects for all indicators and active covariates are unrestricted..
- Variables with effects restricted to zero are still included in the calculation of the overall model statistics such as L^2 .
- Restriction of the set of gamma effects to zero for a covariate causes that covariate to be inactive in the measurement of the latent variable. Setting to zero the effects of **all** covariates causes all covariates to be inactive in which case the estimates for the beta parameters will be identical to the estimates obtained if the covariates were excluded from the model, although the overall model L^2 and related statistics will differ. (Although all covariate effects are set to zero, covariates specified as active are still used to form the overall multiway Table and hence affect the computation of the L^2 statistic. On the other hand, covariates with scale type 'Inactive' affect neither the parameter estimates nor the statistics such as L^2 . The choice as to whether to treat covariates as active (the default) or inactive is a matter of user preference).



For GClasses and GCFactor effects:

Make sure that the name 'GClasses' or of the GCFactor concerned (e.g., 'GCFactor1') is highlighted in the left-most box

- Adding a checkmark for selected indicators removes the default zero restriction on the beta/lambda effects
- Removing a checkmark for selected group covariates, restricts a set of gamma effects to zero (GClasses only).

IMPOSING CLASS INDEPENDENT RESTRICTIONS

Error variance and error covariance parameters are estimated for continuous indicators only, the latter occurring only when direct effects between 2 or more continuous indicators are included in a model. Following model estimation, these parameter estimates (sigmas) are shown in the Parameters Output. In addition, error correlation parameters may be viewed in the 'Error Correlation' subcategory of the Parameters Output. By default, separate estimates for these parameters are estimated for each cluster.

The 'Cluster Independent' box can be used to impose the following class independence restrictions:

Cluster Independent Error Variances (applies to all continuous indicators)

Selection of this option indicates that the error variances are restricted to be equal across classes (class independent) for all continuous indicators.

- ▷ To select this option, click in the check-box preceding 'Error Variances' (cluster independent check on).

Note: This option is only available for Cluster and Regression models (DFactor model variances are always class independent).

Cluster Independent Error Covariances (applies to pairs of continuous indicators for which direct effects have been included in the model - see Residuals Tab below)

This option indicates that the error covariances are restricted to be equal across classes (class independent).

- ▷ To select this option, click in the check-box preceding 'Error Covariances' (cluster independent check on).

Note: This option is only available for Cluster models.



Cluster Independent CFactor effects (default).

By default, any CFactors that have been specified in the Advanced Tab are restricted to be the same across clusters (cluster independent check on). To allow separate CFactor effects for each cluster click to remove the check-mark (cluster independent check off).

ESTIMATING ORDER-RESTRICTED LATENT CLASSES

By default, clusters are not order-restricted. Click this to indicate that the beta parameter effects and cluster-specific means of the indicators (reported in the Profile Output) should be restricted to be monotonically increasing. With nominal indicators this means that the local odds are assumed to be increasing. This option yields what is called an order-restricted LC model.

EQUAL EFFECTS ACROSS INDICATORS

Make sure that the name 'Clusters' (or one of the cluster names) is highlighted in the left-most box. A right-click activates a popup menu that can be used to set the "Equal Effects" option. This option causes all included beta effects to be equal across indicators of the same scale type (nominal indicators should also have equal numbers of categories).



Equal effects for CFactors, GClasses, and GCFactors work in the same way as for Clusters.

DFACTOR MODEL

The DFactor Model Tab allows you to change the number of discrete factors (DFactors) and/or DFactor levels, and/or to restrict selected DFactor effects to zero.

The Factors Name List Box appears in the left-most portion of the Model Tab. It contains names for each DFactor specified in Step 6 ('DFactor1', 'DFactor2', etc.), followed by a name for each level of that DFactor. By default, each DFactor has 2 levels ('Level1', 'Level2').



If any CFactor effects have been specified in the Advanced Tab, specific names for each continuous factor ('CFactor1', 'CFactor2', 'CFactor3') appears at the bottom of this list. In multilevel models, the list also contains GClasses and/or GCFactors.

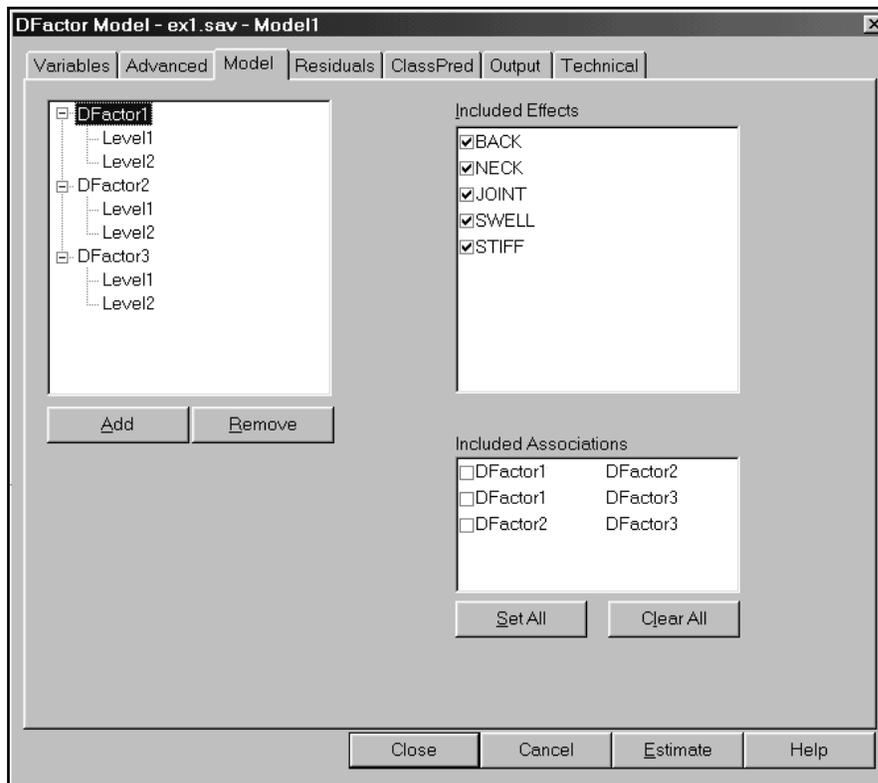


Figure 5-10. Model Tab for DFactor Model

CHANGING THE NUMBER OF DFACTORS OR DFACTOR LEVELS

To increase or decrease the number of DFactors, highlight a DFactor name (e.g., 'DFactor1') in the DFactors Name List Box and click Add to increase or Remove to decrease the number of DFactors. This option works the same as the DFactors box in the Variables Tab.

To change the number of levels of a DFactor, highlight a level name (e.g., 'Level1') and click Add to increase or Remove to decrease the number of levels. Each DFactor can be set to have any number between 2 and 20 levels.

After a DFactor model has been estimated, the number of levels for each DFactor is displayed in the first row of the model summary output file. Note: For each DFactor, the number of levels can be between 2 and 20.

Editing DFactor Names



To customize the DFactor name for your model:

- ▷ Single click the DFactor name (for example, 'DFactor1') in the DFactors Name List Box to highlight it.
- ▷ Single click it again to edit the name (the name will be in an edit box).

Once you have an edit box open, you may also right click to open a menu that allows you to Cut, Copy, Paste, and Delete the DFactor name. This menu also allows you to Undo your last edit or select the whole name for editing (Select All).

INCLUDED EFFECTS BOX

The Included Effects box contains all indicators and active covariates that are included in the current model. Variable names for the active covariates are followed by the characters '<C>'. By default, each of these variable names is preceded by a check (check box equals on) to indicate that for the selected DFactor, beta effects will be estimated for the associated indicator(s) and gamma effects will be estimated for the associated covariate(s). To restrict any of these effects to zero for a DFactor, select the DFactor from the Name List Box and click in the desired check box in the Included Effects box. The check is removed (check box equals off) to indicate that the associated effect is set to zero.



If 1 or more CFactors is included in a model, names for these CFactors ('CFactor1', 'CFactor2', 'CFactor3') appear beneath the DFactor names and DFactor level names. By default, for each CFactor, lambda effects will be estimated for all indicators included in the model. For a selected CFactor, click on the name of that CFactor and the Included Effects box shows the lambda effects to be estimated for that CFactor. To restrict any of these lambda effect(s) to zero, click in the check-box of the indicator(s) and the check is removed. Note: Since CFactor (lambda) effects are estimated only for indicators, the check-box for any Covariates is inactive for CFactor effects.



If GClasses or 1 or more CFactors is included in a model, names for these appear beneath the cluster names. By default, for GClasses and each GCFactor, beta(g) and lambda(g) effects will be excluded for all indicators included. For GClasses or a selected CFactor, click on its name and the Included Effects box shows the effects to be estimated for that latent variable. To add any of these beta/lambda effect(s), click in the check-box of the indicator(s) and the check is added. Note: Since it is not possible to specify regression models for GCFactors, the check-box for any Covariates is inactive for GCFactor effects. Since Group Covariates can be used as predictors in the regression model for the GClasses, the check boxes for the corresponding gamma(g) effects are active and by default on (effect included).

For DFactor effects:

When 2 or more DFactors are included in a model, make sure that the name of the DFactor (e.g., 'DFactor1') is highlighted in the DFactor Name List Box

- Removing a checkmark for selected indicators, restricts the associated beta effects to zero.
- Removing a checkmark for selected active covariates, restricts the associated gamma effects to zero.
- By default, effects for all indicators and active covariates are unrestricted.
- Variables with effects restricted to zero are still included in the calculation of the overall model statistics such as L^2 .
- Restriction of the set of gamma effects to zero for a covariate causes that covariate to be inactive in the measurement of the latent variable. Setting to zero the effects of all covariates causes all covariates to be inactive in which case the estimates for the beta parameters will be identical to the estimates obtained if the covariates were excluded from the model, although the overall model L^2 and related statistics will differ. (Although all covariate effects are set to zero, covariates specified as active are still used to form the overall multiway table and hence affect the computation of the L^2 statistic. On the other hand, covariates with scale type 'Inactive' affect neither the parameter estimates nor the statistics such as L^2 . The choice as to whether to treat covariates as active (the default) or inactive is a matter of user preference).



For CFactor effects:

When 2 or more CFactors are included in a model, make sure that the name of the appropriate CFactor (e.g., 'CFactor1') is highlighted in the left-most box

- Removing a checkmark for selected indicators restricts the corresponding lambda effects to zero.
- The check-box is inactive for covariates.
- By default, effects for all indicators are unrestricted for all CFactors.



For GClasses and GCFactor effects:

Make sure that the name 'GClasses' or of the GCFactor (e.g., 'GCFactor1') is highlighted in the left-most box

- Adding a checkmark for selected indicators removes the default zero restriction on the beta/lambda effects
- Removing a checkmark for selected group covariates, restricts a set of gamma effects to zero (GClasses only).

INCLUDED ASSOCIATIONS BOX

Included Associations refer to the associations between DFactors. By default, the DFactors are assumed to be uncorrelated (DFactor Association Check-box equals off). To include one or more factor association parameters in the model, click in the associated DFactor Association Check-box.

EQUAL EFFECTS ACROSS INDICATORS

When 2 or more DFactors are included in a model, make sure that the name of the DFactor (e.g., 'DFactor1') is highlighted in the DFactor Name List Box. A right-click activates a popup menu that can be used to set the "Equal Effects" option. This option causes all included beta effects to be equal across indicators of the same scale type (nominal indicators should also have equal numbers of categories).



Equal effects for CFactors, GClasses, and GCFactors works in the same as for DFactors.

LC REGRESSION MODEL

Various restrictions are available for intercepts and predictor effects. In addition, for models with continuous dependent variables, restrictions are available for error variances. The various restrictions include class independent restrictions, order restrictions, zero restrictions, fixed-value (offset) restrictions, and equality restrictions. By default, no restrictions are imposed.

Restrictions can be placed prior to estimating an initial model to incorporate prior knowledge. For example, the zero constraints make it possible to specify a different regression model -- with different predictors -- for each latent class based on a priori knowledge about the classes. A specific application for this is a model with a random responders class for which all predictor effects are zero. An application of the equality constraints is the possibility of defining a DFactor-like structure in which, for example, one DFactor influences the intercept and another the predictor effects. Ordering constraints are important if one has a priori knowledge on the direction of an effect. For example, the price effect on a product rating is usually assumed to be negative (or non-positive) in each latent class (segment). The estimates obtained with Regression will be constrained to be in agreement with this assumption if the price effect is specified to be **Descending**.

Alternatively, restrictions may be employed post-hoc to estimate a new model after viewing the results from an estimated model.

For Regression, when **K** classes have been specified in Step 6, the Model Tab contains **K** individual columns for each of the **K** latent classes (1,2, ...,K). Additional columns are labeled **Class Independent** and **Order Restriction**. When the range option is used in Step 6 to specify the estimation of models with different numbers of latent classes, class-specific columns are absent from the Model Tab and only the class independent and order restrictions can be applied. Such restrictions are applied to each of the models generated by the range option.

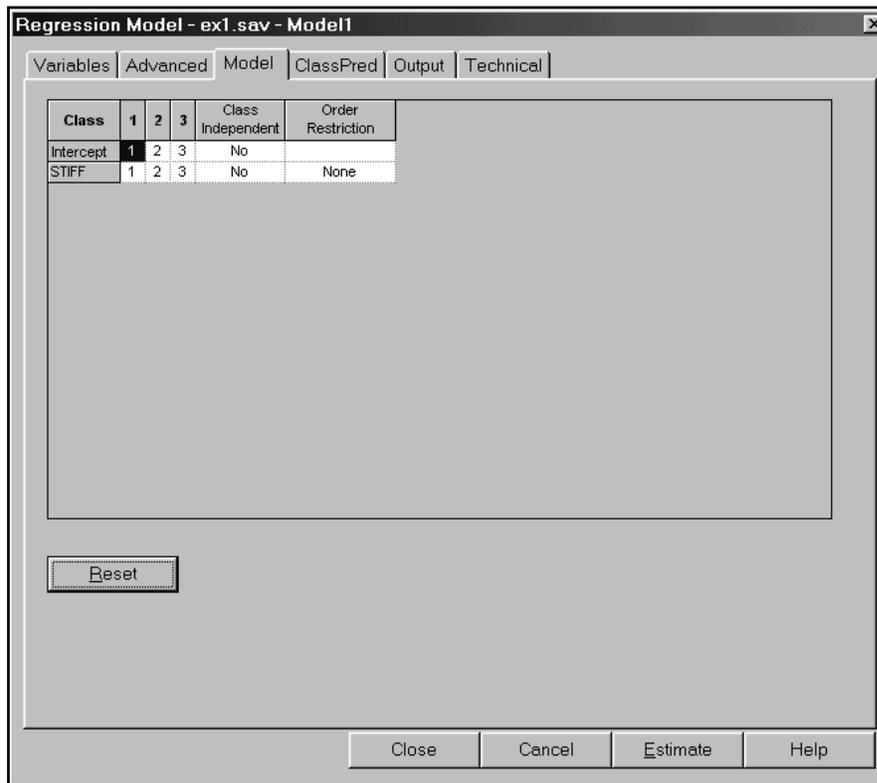


Figure 5-11. Model Tab for LC Regression Model.

Selected intercept and predictor effects can be set equal to zero for certain classes (No Effect), equated across two or more classes (Merge Effect), or equated across all classes (Class Independent). The effects of a predictor can be restricted to be ordered in each class (Ascending or Descending). Error variances can be equated between selected classes (Merge Effect) or between all classes (Class Independent). The effect of a numeric predictor on a dependent variable that has a scale type other than nominal can be fixed to 1 in one or more classes (Offset).

The first row consists of the label 'Intercept', followed by separate row for each predictor. Restrictions can be set separately for each row. By default, no restrictions are imposed. This is indicated by the unique integers (1,2, ...,K) that appear in each row, by the label 'No' indicating that the class independence restriction has not been imposed and the label 'None' indicating that no order restrictions have been selected for any predictors.

SPECIFYING EQUALITY RESTRICTIONS ACROSS CLASSES

The effects for a predictor may be specified as Class Independent (Fixed Effects) or Class Dependent (Random Effects, default). When the Class Independent restriction is applied, the regression coefficient for that predictor is restricted to be equal between each of the K latent classes (segments). When Class Independent is selected for a predictor, an '=' appears to the right of the predictor name in the Variables Tab. For a 1-class regression model (K=1), selection of this option has no effect.



To select this option

- ▷ Right click on the desired cell(s) in the column labeled 'Class Independent' to retrieve the popup menu.

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

- ▷ Select 'Yes'. The 'No' changes to 'Yes' in the selected cells, and the indices in the selected rows all change to '1' to indicate that the effects for classes 2, 3, etc. are all restricted to be equal to the corresponding effects for class 1.

Alternatively, the restriction of class independence can be imposed as follows:

- ▷ Select all the cells containing the indices for a chosen row
- ▷ Right click to bring up the pop-up menu
- ▷ Select 'Merge Effect'.
- ▷ To undo these restrictions, reselect the cells, right click and select 'Separate Effect'.

This alternative way of imposing the *class independence* restriction can also be used to specify equal effects between some but not all the classes. Simply select those class indices for which the across-class equality restriction is desired and select 'Merge Effect'. The indices selected will change to be equal. 'Merge Effect' can be used more than once for a given row, so that it is possible to specify that certain effects for classes 1 and 2 be restricted to be equal and that the effects for classes 3 and 4 be equal. After imposing these restrictions, the associated indices will appear as '1 1 3 3'.

Note: the numbers indicate to which class the effect for the class concerned is equated.

The regression intercept may also be restricted to be class independent. To set this option, right click on the row labeled 'Intercept' in the column labeled 'Class Independent' to retrieve the popup menu and select 'Yes'.

Note: For dependent variables having scale type 'Ordinal', a third option 'No - Simple' is also available. Rather than *complete* class independence ('Yes'), a 'simple' implementation of class independence analogous to a class independent intercept specification in the linear regression model (continuous scale type) is used where the (partial) dependent variable means are taken to be class independent rather than the entire marginal dependent variable distribution. For further details of this 'class-independent simple' specification for the intercept in ordinal regression models, see section 5.3 of Technical Guide

SPECIFYING ZERO RESTRICTIONS

To restrict one or more effects to zero, select the desired effects, right click, and select 'No Effect'. This causes a '-' to appear in the selected cells which indicates that the effect(s) associated with these cells is now restricted to zero. These menu options can also be used in combination with each other to produce a desired result. For example, first selecting *class independence*, followed by selecting 'Separate Effect' in one of these cells causes the index for the selected cell to return to its default setting. However, the remaining cells in that row remain restricted to be equal and 'Yes' automatically changes to 'No' in the *Class Independent* column.

SPECIFYING ORDER RESTRICTIONS

Order Restriction can be used to indicate that the regression coefficient is monotonically increasing (Ascending) or decreasing (Descending).

To impose an *ordering* restriction, right click on one or more cells containing a 'None' label and right click to retrieve the relevant pop-up menu. Select either *Ascending* or *Descending* to impose the desired ordering restriction on the selected predictor effect(s). This restriction is imposed across all classes. With a dependent variable of a scale type other than 'nominal', the usage and interpretation of this restriction is straightforward. In the case of a nominal dependent variable, the difference between parameters of adjacent categories of the dependent variable will be in agreement with the specified order restrictions (see section 5.3 of Technical Guide).

SPECIFYING FIXED-VALUE RESTRICTIONS

With the *Offset* option one can indicate that the effect of a numeric predictor should be fixed to 1 in selected classes. The option is not available for Nominal dependent variables.

Note: It is possible to use the offset restriction to fix the effect to non-zero quantities other than 1 as well. For example, to restrict the effect of a numeric predictor to equal some desired constant c , you would first rescale that predictor by multiplying it by c . You would then use the rescaled version of the predictor in the model.

A specific application occurs in the case in which one would like to fix the response probability to 0 for certain predictor values for certain classes. Suppose that the dependent variable is 1='buy', 0='no buy' and that you suspected that one latent class existed that always responded 'no buy' when the predictor variables reflected a certain pattern (e.g., PRICE \$50). This option could be used to identify this latent class by creating a dummy predictor variable, say coded [-100, 0] where the value -100 refers to a situation which is suspected to always result in a 'no buy' for some class (fixing a logit coefficient to -100 amounts to fixing the probability of a 1='buy' response to 0).



INCLUDING CFACTOR, GCLASSES, AND GCFATOR EFFECTS IN THE MODEL FOR THE DEPENDENT VARIABLE

When included, CFactor, GClasses, and GCFactors appear as additional columns in the Model Tab.

Check box switches can be added or removed to indicate whether CFactors, GClasses, and/or GCFactors affect the intercept and/or the predictor coefficients,

By default:

- CFactor1: intercept checked and other terms unchecked.
- CFactor2 and CFactor3: all effects checked.
- GClasses and GCFactors: all effects unchecked.

Note: For all included effects the default is "Class Independent = Yes". This can be changed using the "Class Independent =No", "Merge Effects", and "No Effect" options.

Applications:

- random intercept regression model (1 CFactor, default specification)
- 3-level LC regression (GClasses)
- 3-level random-effects modeling (GCFactors combined with CFactors).

ClassPred Tab: Restricting Cases Known (Not) to Belong to a Certain Class or Classes

With this option one can specify that one or more specific cases can belong to a certain class or certain classes only. To use this feature, select a variable from the list box in the ClassPred Tab to be used as the Known Class Indicator and click Known Class. The variable moves to the Known Class Indicator Box and the Assignment Table becomes active. For each category of the Known Class indicator you then specify to which classes the cases with that category code may belong (*or not belong*) using the Assignment Table. For example, Figure 5-12 illustrates a 4-Cluster model (4 columns) where the variable 'classind' is used as the Known Class Indicator. Cases for which 'classind=1' are allowed to be in cluster 1 only; those for which 'classind=2' are allowed to be in cluster 2 only; all other cases ('classind=3') may be assigned to any of the 4 clusters.

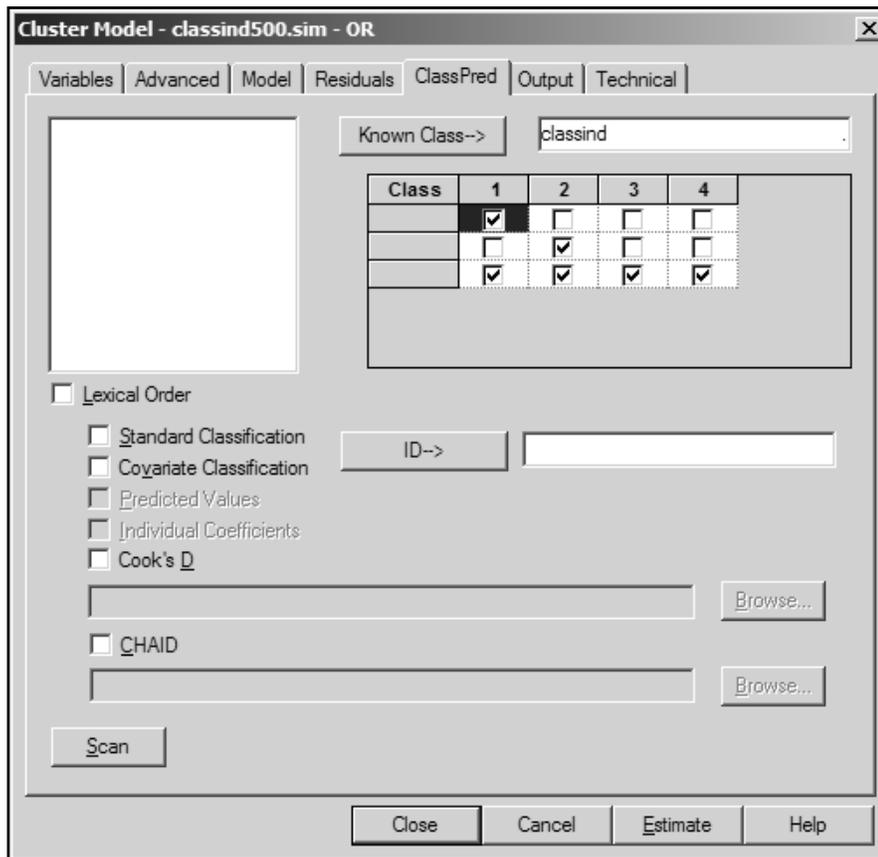


Figure 5-12. ClassPred Tab for Cluster Model with a Known Class Indicator

This option is useful if you have a priori class membership information for some cases (pre-assigned or pre-classified cases) or if membership to certain classes is very implausible for some combinations of observed scores.

KNOWN CLASS - CLASS INDICATOR

In applications where a subset of the cases are known with certainty not to belong to a particular class, or particular classes, you can take advantage of this information to restrict their posterior membership probability to 0 for one or more classes and hence classify these cases into one of the remaining class(es) with a total probability = 1. This feature allows more control over the segment definitions to ensure that the resulting classes

are most meaningful. Common applications include:

- 1) using new data to refine old segmentation models while maintaining the segment classifications of the original sample
- 2) archetypal analysis - define class membership a priori based on extreme response patterns that reflect theoretical "archetypes"
- 3) partial classification -- high cost (or other factors) may preclude all but a small sample of cases from being classified with certainty. These cases can be assigned to their respective classes with 100% certainty, and the remaining would be classified by the LC model in the usual way
- 4) certain cases may be known to be "type 1 OR type 2" (e.g., 'clinically depressed' or 'troubled'). By excluding such cases from being in say class 3 = 'healthy', such cases can be pre-assigned to be in class 1 or 2, while additional cases may be freely classified into any class
- 5) post-hoc refinement of class assignment where modal assignment for certain cases is judged to be implausible based on the desired interpretation of the classes.

In addition, this option may also be used to specify multiple group models by including the group variable as both a Known Class Indicator and as an active covariate. For further details of this, see section 2.5 of Technical Guide.

Note: The Known Class option is not available in cluster and regression if the Range option has been used in the Variables Tab.

For DFactor models, this option applies only to levels of DFactor 1.



To select known classes (clusters/classes/DFactor1 levels):

- ▷ Select one variable from those appearing in Variables List Box (located in the upper left-hand portion of the ClassPred Tab). Variables appearing here are those that have not been previously selected as Indicators or Covariates.
- ▷ Click Known Class to move that variable to the Known Class Box and the class assignment window beneath the Known Class Box becomes active.
- ▷ A separate row appears for each category/code/value taken on by the known class indicator
A separate column appears for each class.
- ▷ Click on the appropriate boxes to select or deselect the possible assignment of the categories to certain classes.

A checkmark off means that the posterior membership probability is restricted to zero for that class for cases in that category of the known class indicator.

By default, the checks are assigned as follows:

For a K-class model, a category with a code of K on the Known Class Indicator is assigned to only class

K. Categories coded less than 1, greater than K or missing are assigned to all classes (i.e., no restrictions). Missing values are not shown in the table.

Note: For the example in Figure 5-12 above, all cases are coded either 1, 2 or 3 on the variable 'classind' (i.e., no missing values). Those coded 'classind=1' and 'classind=2' are maintained at their default specifications on the table, while the default specification for cases coded 'classind=3' was changed (from 'cluster 3 only' to 'any cluster' -- all 4 cluster columns checked). This specification would be obtained by default if those coded '3' on the classind variable were instead coded as 'missing'. In this situation, the table would differ from that shown in Figure 5-12, in that the 3rd row of the table would not appear, since that category would be coded 'missing'.

For further information, see section 2.5 of the Technical Guide.

Residuals Tab: Including Direct Effects in Model (**CLUSTER AND DFACOR MODELS**)

From the Residuals Tab you may specify a direct effect to include in a model. All pairs of variables eligible for a direct effect parameter appear. To include a direct effect, click in the check-box and a check appears. Direct effect parameters will be estimated for the pairs of variables that have been so selected (direct effect check-box equals on). The inclusion of direct effects is one way to relax the assumption of local independence.

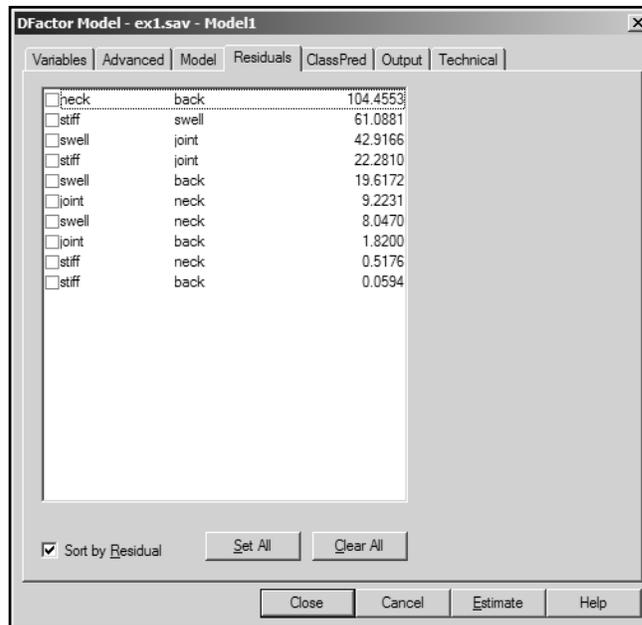


Figure 5-13. Residuals Tab for DFactor Model

Direct effects may be specified prior to estimating a model or post-hoc after examining the results of a model. To assist you in choosing pairs of variables for which direct effects may improve the fit of a model, bivariate residuals (BVR) may be output.

Bivariate residuals are listed in the Bivariate Residuals section of the Outline pane when this option has been selected in the Output Tab. They are useful for diagnostic purposes. Advanced users may wish to include one or more Direct Effects Parameters that are associated with large bivariate residuals in a model.

A direct effect between two indicators and direct effects of selected covariates on selected indicators can be

selected for inclusion in a model using the Residuals Tab on the Analysis dialog box. After estimation, these effects will be listed in the Parameters output in a section labeled 'Direct Effect Parameters'.

Note: If this is your first model for a data file, once you have specified the model in the Variables Tab, all pairs of indicators and covariates will be listed without any bivariate residuals in the Residuals Tab.

After you have estimated a model, the Residuals Tab will also contain a bivariate residual associated with each direct effect. Reopen the Analysis dialog box and now the Residuals Tab contains the residuals based on the model sorted from high to low according to the magnitude of the residual.

The example in Figure 5-13 shows that the direct effect between the variable pair NECK and BACK is associated with the largest residual is. This residual value (chi-square divided by its degrees of freedom) is 104.4553, which is much larger than 1, the reference value for these residuals.

In situations where there is only one large diagnostic value, a new model can be estimated by selecting the variable pair (a pair is selected when a checkmark appears in the box to the left), and clicking Estimate, thus adding the associated direct effect parameter to the current model.

Note: If there are several large residuals (as in Figure 5-13), a common strategy is to include the corresponding direct effects one at a time as needed, each time re-estimating the model, and checking the updated residuals after each new model is estimated before including additional direct effects. This is because once you have included a direct effect in a model, all of the residuals in that new model may be small.

After estimating, the bivariate residuals will be listed as 0 (or very close to 0) for any variable pair for which a parameter has been included in the model. Also note that if you estimate a new model that included residuals, Latent GOLD will retain all of your settings when estimating additional models, including the residuals selected for inclusion in the prior model.

Check boxes:

Sort by Residual. By default, bivariate residuals are sorted from largest to smallest (sort residual check box equals on). If this option is not checked, the residuals will be listed in default order which is based on the order the variables were entered into the model in the Variables Tab.

Set all. Click this to include direct effects for all eligible pairs of variables.



Advanced Tab (REQUIRES THE ADVANCED VERSION OF LATENT GOLD)

The Advanced Tab is divided into 3 areas according to the section labels 'Survey', 'Continuous Factors', and 'Multilevel Model'.

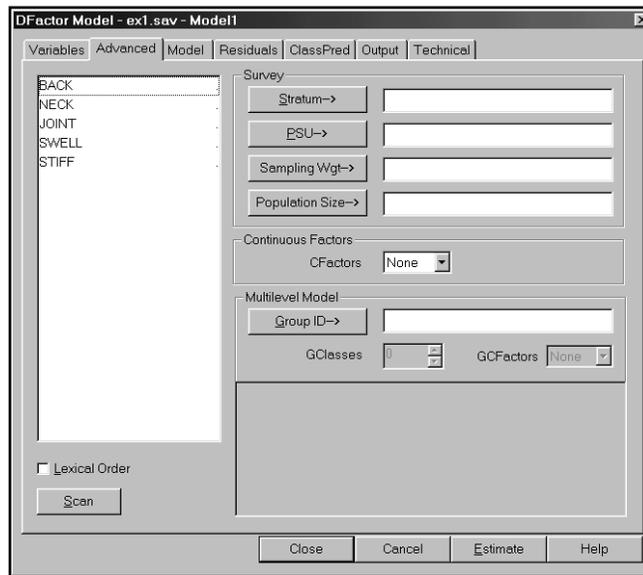


Figure 5-14. Advanced Tab for DFactor Model

The variables displayed in the variable list (left-most box of the Advanced Tab) are those that have not been specified previously for use as an indicator, dependent variable, predictor, covariate, known class indicator, case ID, case weight, or replication weight. These variables are eligible for use with any of these 3 advanced options.

SURVEY

This advanced option can be used to specify information on the sampling design that was used to obtain your data. The program computes the design effect, as well as reports sampling design corrected standard errors and Wald statistics. Four aspects of the sampling design can be taken into account: stratification (Stratum), clustering (PSU), weighting (Sampling Wgt), and finite population size (Population Size).

For more details, see section 11 of Technical Guide.

Stratum

The Stratum variable specifies the stratum to which a case belongs. When no Stratum variable is specified, it is assumed that all cases belong to the same stratum; that is, that there is only one stratum.

PSU

The PSU(Primary Sampling Unit) variable is used for two-stage cluster samples. It specifies the (sampling) cluster to which a case belongs. PSUs are assumed to be nested within strata. When no PSU variable is specified, it is assumed that each case from a separate PSU.

Sampling Wgt

The Sampling Wgt variable contains a sampling weight.

Rescale (default) vs. No Rescale. Upon selecting a variable to be used as a Sampling Wgt, the symbol <R> appears in the Sampling Wgt box to the right of the variable name to indicate that the weights will be rescaled. Rescaling of the original weights are accomplished by multiplying them by a constant

such that the sum of the weights equals the sample size.

- ▷ Right click on the variable name and select 'No Rescale' from the popup menu to maintain the weights without rescaling. Upon selection of 'No Rescale', the <R> symbol is removed.

Active (default) vs. Inactive. By default the sampling weights are used in the estimation to compute pseudo maximum likelihood estimates as indicated in Section 11.1 of Technical Guide. If the sampling weight were instead specified as a Case Weight in the Variables Tab, the resulting parameter estimates would be the same as when the Active option is used here but the standard errors are not correct.

The inactive option for sampling wgt employs an alternative 2-step estimation algorithm developed by Vermunt and Magidson (2001).

- ▷ Right click on the variable name and select 'Inactive' from the popup menu to select this option. Upon selection of 'Inactive', the <I> symbol appears in the Sampling Wgt box to the right of the variable name.

If the sampling weight variable were instead not used at all in the estimation (not specified as either a Case Weight nor a Sampling Wgt), the parameter estimates obtained would be the same as when the Inactive option is used here, but the sizes of the latent classes would be biased. The advantage of this method over the Active option is that the unweighted estimates may be more stable. See Section 11.2 of the Technical Guide for further information about these options.

Population Size

The Population Size variable can be used to specify either the size of the population (# of PSUs in the Stratum concerned) or the population fraction. The variable is assumed to be a population fraction when it is smaller or equal to 1. This option can be use for finite population corrections.

Continuous Factors

This advanced option can be used to include up to 3 continuous latent variables (CFactors) in an LC Cluster, DFactor or Regression model. In the Cluster and DFactor Modules, the use of CFactors yields (mixture variants of) factor-analytic models and various types of latent trait or IRT models. In the Regression (and Choice) Modules, it yields models with continuous random effects.

By default, the CFactors box is set to 'None'. To include CFactors in a model, click to open the drop down menu and select the number of CFactors to include in the model (1, 2, or 3). When 1 or more CFactors are included in the model, they appear on the Model Tab for further model specification.

By default, CFactors use 10 nodes to approximate normally distributed variables. To improve precision of the estimates, the number of nodes may be increased to a value as high as 50 (or reduced as low as 2). This change is made in the Continuous Factors section of the Technical Tab (see Step 8 for details).

WARNING: Inclusion of CFactors in a model may substantially increase the amount of time required to estimate the model. For example, inclusion of 2 CFactors, results in $10 \times 10 = 100$ nodes used to approximate the bivariate normal distribution for these CFactors. Increasing the number of nodes to 50 results in $50 \times 50 = 2500$ nodes, which will substantially increase the

amount of estimation time. For further details, see section 9.1 of Technical Guide.

When used with a 1-class cluster or 1-class regression model, the result is not a LC model. For example, a 2-CFactor 1-class model with continuous indicators is identical to a traditional linear factor analysis (FA) model. When one or more indicators is other than continuous, this model becomes an IRT variant of the FA model. For further details regarding the various kinds of applications with CFactors, see Section 9.2 of Technical Guide.

MULTILEVEL MODEL

This advanced option is used to specify a multilevel extension to an LC Cluster, DFactor or LC Regression model which allows for explanation of the heterogeneity not only at the case level, but also at the group level.

Heterogeneity at the group level is explained by the inclusion of group-level classes (GClasses) and/or group-level CFactors (GCFactors) in a model.

Group ID

The Group ID variable indicates to which higher-level unit or group each case belongs. Upon selecting a variable as the Group ID, the Group Specification Box in the lower-right portion of the Advanced Tab is activated.

GClasses

This option assumes that groups belong to one of a set of latent classes of groups, the number of which is specified with GClasses (Group-level Classes). This yields the *nonparametric* variant of the multilevel LC model. By default, the GClass box is set to 1. To use this option to specify 2 or more GClasses

Click the up arrow in the drop-down box to increase the number of GClasses to 2 or more (up to 100). The GCFactors then appear in the Group Specification Box below.

GCFactors

This option assumes that groups differ with respect to their scores on one or more group-level continuous factors (GCFactors) or group-level random effects. This yields the *parametric* variant of the multilevel LC model. Click on the drop-down box to select the number of GCFactors. The GCFactors then appear in the Group Specification Box below.

GClasses and GCFactors may both be specified to combine the parametric and nonparametric approaches.

GClasses and GCFactors may

- affect the intercept and the covariate effects in the regression model for the Clusters, DFactors or Classes
- have direct effects on the indicators (see Model Tab in Cluster/DFactor)
- affect the intercept and the predictor effects in the model for the dependent variable (see Model Tab in Regression)

GClasses may themselves be affected by Group-level covariates (GCovariates).

When CGClasses or CGFactors are included in the model, they appear on the Model Tab for further model specification as described earlier; that is, to include effects on the indicators and the dependent variable. To include

GClass and/or GCFactor effects in the regression model for the Clusters, DFactors or Classes, use the Group Specification Box.

Group specification box

The Group specification box at the lower-right of the Advanced Tab contains a column for GClasses and additional columns for each GCFactor specified. Click in the check boxes to allow estimation of desired parameters.

When GClasses and/or GCFactors are included, it is assumed that these affect the intercept in the regression model for the Clusters, Classes, or DFactors. This yields the standard multilevel latent class model in which class sizes are assumed to differ across groups by using a (parametric or nonparametric) random-intercept model for the latent classes.

GClasses and GCFactors may also be allowed to affect the covariate effects in the regression model for the Clusters, Classes, or DFactors. This is accomplished by checking the corresponding terms on the Advanced Tab.

GClasses play a role similar to the one of the Classes in a LC regression model: effects can be GClass independent (check off) or GClass dependent (check on). GCFactors play a role similar to CFactors in a random-effects regression model: effect can be assumed to be fixed (check off) or random (check on).

See Section 10 of Technical Guide details and application types for multilevel models.

Step 8: Set Technical Options (Technical Tab)

The Technical Tab of the Analysis dialog box contains various technical options that are available to control the estimation process and need to be set before a model is estimated to take effect. In addition to the usual kinds of options affecting the number of iterations and convergence limits, the other options - Start Values, Bayes Constants, and treatment of missing data can affect the quality of the output results. For detailed descriptions of these options, see sections 6.2-6.7 of Technical Guide.

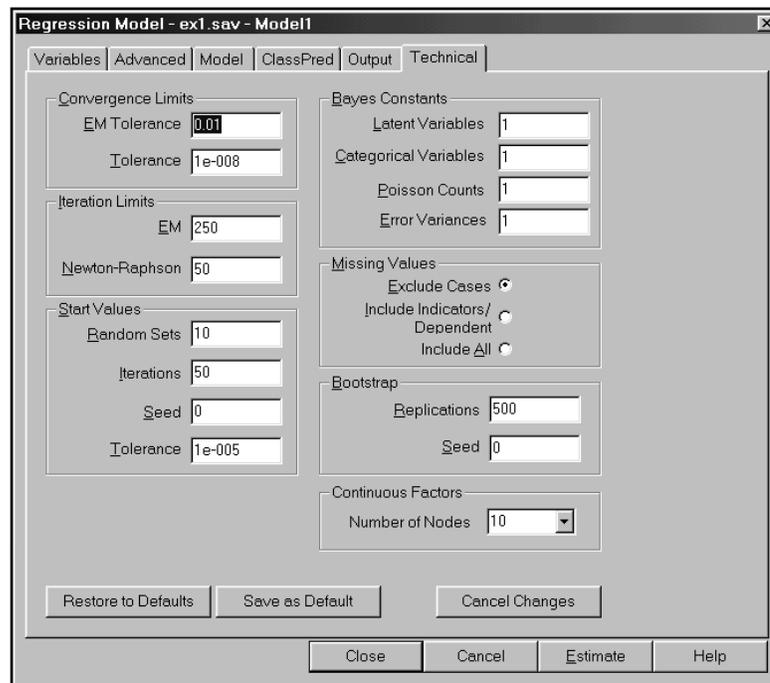


Figure 5-15. Technical Tab for Regression Model

CONVERGENCE LIMITS

EM Tolerance

EM Tolerance is the sum of absolute relative changes of parameter values in a single iteration. It determines when the program switches from EM to Newton-Raphson (if the NR iteration limit has been set to > 0). Increasing the EM Tolerance will switch faster from EM to NR. To change this option, double click the value to highlight it, then type in a new value. You may enter any non-negative real number. The default is 0.01. Values between 0.01 and 0.1 (1% and 10%) are reasonable.

Tolerance

Tolerance is the sum of absolute relative changes of parameter values in a single iteration. It determines when the program stops its iteration. The default is 1.0×10^{-8} which specifies a tight convergence criterion. To change this option, double click the value to highlight it, then type in a new value. You may enter any non-negative real number.

ITERATION LIMITS

EM Iterations

Maximum number of EM iterations. The default is 250. If the model does not converge after 250 iterations, this value should be increased. You also may want to increase this value if you set NR iterations = 0. To change this option, double click the value to highlight it, then type in a new value. You may enter any non-negative integer.

Newton-Raphson

Maximum number of NR iterations. The default is 50. If the model does not converge after 50 iterations, this value should be increased. To change this option, double click the value to highlight it, then type in a new value. You may enter any non-negative integer. A value of 0 is entered to direct Latent GOLD to use only EM, which may produce faster convergence in models with many parameters or in models that contain continuous indicators or dependent variables.

START VALUES

To reduce the likelihood of obtaining a local solution, the following options can be used to either increasing the number of start sets, the number of iterations per set, or both.

Random Sets

The default is 10 for the number of random sets of starting values to be used to start the iterative estimation algorithm. Decreasing the number of sets of random starting values for the model parameters reduces the likelihood of converging to a local (rather than global) solution. To change this option, double click the value to highlight it, then type in a new value. You may enter any non-negative integer. Using either the value 0 or 1 results in the use of a single set of starting values.

Iterations

This option allows specification of the number of iterations to be performed per set of start values. Latent GOLD first performs this number of iterations within each set and subsequently twice this number within the best 10% of the start sets. For some models, many more than 20 iterations per set may need to be performed to avoid local solutions.

Seed

The default value of 0 means that the Seed is obtained during estimation using a pseudo random number generator based on clock time. Specifying a non-negative integer different from 0, yields the same result each time.

If the current model setup was obtained by opening an .lgf file associated with a previously estimated model, 1) the Seed will not be 0 but will be the Best Start Seed for that model as specified in the .lgf file, and 2) the Random Sets parameter will be set equal to 0. This procedure assures that the model estimated is exactly the same model obtained when originally estimated, as long as the .lgf file was created using Latent GOLD 4.0 (see Warning below).

To specify a particular numeric seed (such as the Best Start Seed reported in the Model Summary Output for a previously estimated model), double click the value to highlight it, then type in (or copy and paste) a non-negative integer. When using the Best Start Seed, be sure to deactivate the Random Sets option (using Random Sets = 0). For further details see section 6.6 of Technical Guide.

WARNING: Due to improvements in this option in Latent GOLD 4.0, the random seed obtained from earlier versions of Latent GOLD will not necessarily reproduce the original model and has an increased chance of resulting in a local solution. Hence, if you open an .lgf file that was created using an earlier version of Latent GOLD, you should make sure to restore the default value of 0 and increase the value for Random Sets to the default value of 10 or some other desired quantity.

Tolerance

Indicates the convergence criterion to be use when running the model of interest with the various start sets. The definition of this tolerance is the same as the one that use used for the EM and Newton-Raphson Iterations.

BAYES CONSTANTS

The Bayes options can be used to eliminate the possibility of obtaining boundary solutions. You may enter any non-negative real value. Separate Bayes constants can be specified for three different situations:

Latent Variables

The default is 1. Increase the value to increase the weight allocated to the Dirichlet prior which is used to prevent the occurrence of boundary zeroes in estimating the latent distribution. The number can be interpreted as a total number of added cases that is equally distributed among the classes (and the covariate patterns). To change this option, double click the value to highlight it, then type in a new value.

Categorical Variables

The default is 1. Increase the value to increase the weight allocated to the Dirichlet prior which is used in estimating multinomial models with variables specified as Ordinal, Nominal or Binomial Count. The number can be interpreted as a total number of added cases to the cells in the models for the indicators (model for dependent) to prevent the occurrence of boundary zeroes. These pseudo cases are divided equally across classes and predictor/covariate patterns, and in accordance with the observed marginal distribution across categories of the indicator (dependent variable) concerned. To

change this option, double click the value to highlight it, then type in a new value.

Poisson Counts

This prior is equivalent to adding a specified number of events to the data without changing the overall Poisson rate. In other words, the number of exposures is adjusted accordingly. This prior prevents boundary solutions in regression models where the Poisson count scale type is used. The default value for this Bayes constant is 1. To change this option, double click the value to highlight it, then type in a new value.

Error Variances

The default is 1. Increase the value to increase the weight allocated to the inverse-Wishart prior which is used in estimating the error variance-covariance matrix in models for continuous dependent variables or indicators. The number can be interpreted as the number of pseudo-cases added to the data, each pseudo-case having a squared error equal to the total variance of the indicator (dependent variable) concerned. Such a prior prevents variances of zero from occurring. To change this option, double click the value to highlight it, then type in a new value.

MISSING VALUES

The Missing Values option allows for the inclusion of records containing missing values on covariates and predictors as well as records containing missing values on the indicators. Including cases with missing values on covariates and predictors causes the mean to be imputed for the scale type numeric and the effect of the missing value category to be equated to zero for the scale type nominal. Missing values on indicators and dependent variables are handled directly in the likelihood function.

Exclude cases

For Regression, selection of this option excludes all replications having missing values on the dependent variable or any of the predictors and all cases having missing values on any of the active covariates. For Cluster/DFactor, selection of this option excludes all cases having missing values on any of the indicators or active covariates.

Include indicators/dependent

For Regression, selection of this option excludes replications having missing values on any of the predictors and cases having missing values on any of the active covariates. For Cluster/DFactor, selection of this option excludes cases having missing values on any of the active covariates. After exclusion of these cases, the remaining cases with missing values on the dependent variable (Regression) or on any indicator (Cluster and DFactor) are included in the analysis and handled directly in the likelihood function.

Include all

Selection of this option includes all cases and replications in the analysis regardless of the presence of missing values. Cases or replications with missing values on the dependent variable (Regression) or on any indicator (Cluster and DFactor) are included in the analysis and handled directly in the likelihood function. Missing values on Predictors (Regression Module), or active covariates (Regression, Cluster and DFactor Modules) are imputed using Latent GOLD's imputation procedure

Inclusion in a model of covariates designated as inactive has no effect on which cases are excluded. Therefore, these missing values options have no effect with respect to the presence or absence of missing values on covariates specified to be inactive

BOOTSTRAP OPTIONS (BOOTSTRAP L², BOOTSTRAP -2LL DIFF)

The Technical Tab contains options for specifying the number of Replications and a Seed for both the Bootstrap L² and the conditional bootstrap (Bootstrap -2LL Diff) procedures. Either of these bootstrap procedures can be requested from the Model Menu for an estimated model as described in Step 10.

Replications

The default for the number of replication samples is 500. In most applications this number is large enough. The program also reports the Monte Carlo standard error of the p-value. By increasing this number, a more precise estimate of the p-value is obtained since the Monte Carlo error is reduced. With large models, to speed up the estimation you may consider reducing the number of replications.

Seed

Seed can be used to specify the seed that is used to generate the replication data sets (the default value 0 means random seed) for either the Bootstrap of L² or the Conditional Bootstrap ('Bootstrap -2LL Diff') procedures. Because of Monte Carlo simulation error, these bootstrap procedures yields a slightly different p-value each time that it is repeated, along with an estimate of the standard error. Specifying a particular seed guarantees the same result each time. By specifying the seed to be equal to the bootstrap seed reported in the Model Summary Output, one can replicate a previous run. In most bootstrap applications one will only use the Replications option.

If the Save Definition option in the File Menu is used to save a .lgf definition file for a model resulting from the Bootstrap, the Bootstrap Seed is saved. To reproduce results obtained from the Bootstrap of L², see the Note in Step 10 in the section on the Bootstrap p-value (L²). To reproduce results obtained from the Conditional Bootstrap, see Note2 in Step 10 in the section on the Conditional Bootstrap (-2LL Diff). For the Conditional Bootstrap, only the Bootstrap Seed associated with the source model is utilized.



ADVANCED. CONTINUOUS FACTORS

Number of Nodes

If 1 or more (group-level) continuous factors are specified in the Advanced Tab, this option determines the number of nodes used to approximate their normal distribution. By default, 10 nodes are used. Decreasing this number (minimum is 2 nodes) will speed up estimation time but reduce precision of the multivariate normality of the CFactors/GCFactors. For further details see Sections 9.1 and 10.1 of Technical Guide.

DEFAULT OPTIONS

Click **Save as Default** to save the current technical settings as the new default values.

Click **Restore to Defaults** to restore the technical options to their last default values.

Click **Cancel Changes** to cancel any changes that have been made to the Technical options and not saved.

Step 9: Set Output Options (Output Tab)

The Output Tab allows you to select various types of output listings which appear following estimation of your model(s). Each additional output selected for a model will be listed in the Outline pane. Chapter 6 provides a detailed specification of each of these output listings.

The ClassPred Tab allows you to output a data file containing selected classification and prediction information. This output file is only available if the input data is either an SPSS .sav file or an ASCII text file.

The Technical Guide provides related technical information, formulae and equations for all parameters and related statistics and other output.

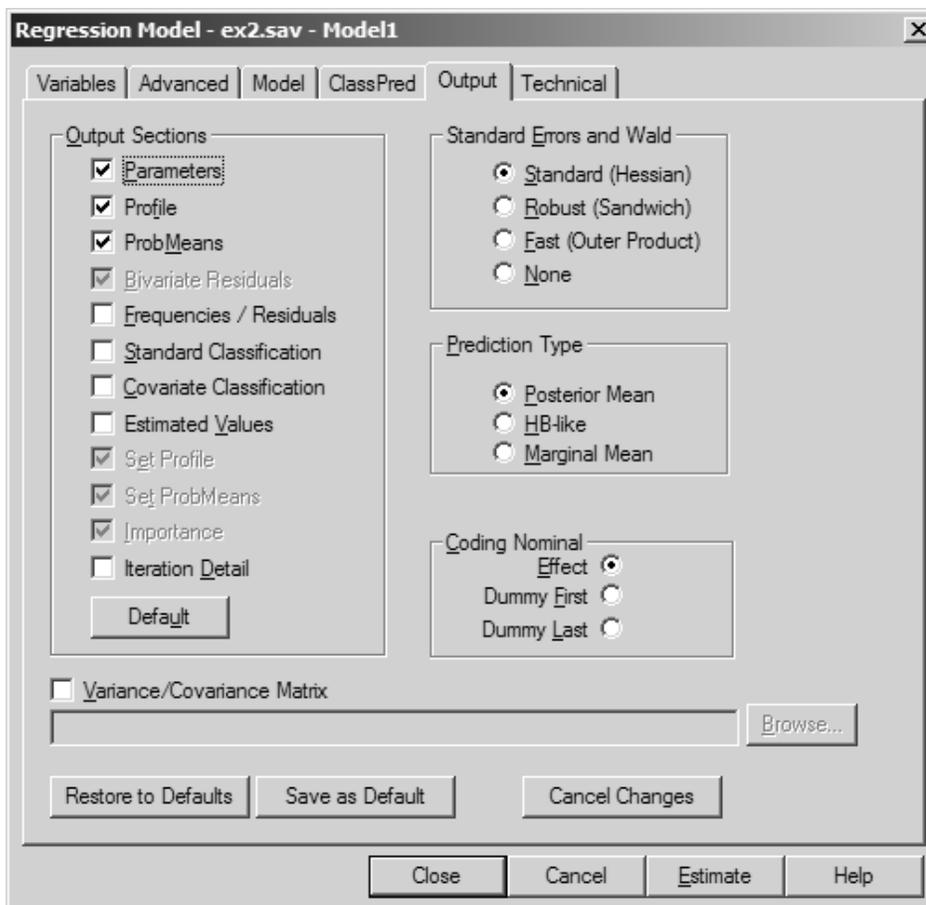


Figure 5-16. Output Tab for LC Regression Model

OUTPUT SECTIONS

A checkmark indicates that the associated Output listings are produced. For details of these output files, see Chapter 6. By default the following are produced (checkmark equals on):

Parameters. Shows/hides Parameters in Output Window

Profile. Shows/hides Profile in Output Window

ProbMeans. Shows/hides ProbMeans in Output Window

They may be de-selected by clicking the check-box (Output check equals off) in which case this type of output will not appear.

The remaining output listings can be obtained by clicking the check-box (Output check equals on).

Bivariate Residuals. Produces a Table containing bivariate residuals. This output is not available with Regression Models. The output file will be listed as 'Bivariate Residuals' in the Outline pane.

Frequencies/Residuals. Produces an output file containing observed and estimated frequencies, and standardized residuals for each combination of variables. This output is not available if any variables in an analysis have been specified as Continuous or Count. It is also not available with Regression Models for which an ID variable has been specified (for repeated measures, for example). This output file will be listed as 'Freqs/Residuals' in the Outline pane.

Classification Output (optional)

Standard Classification. Produces an output file listing containing posterior membership probabilities and other information used to classify cases into the appropriate cluster, latent class or DFactor level. This output file will be listed as 'Standard Classification' in the Outline pane.

Each row in the Standard Classification output corresponds to a distinct observed data pattern in the data file. For Cluster and Regression models, each of these rows contains the estimated probability of being in each class. For DFactor models, each row in the Standard Classification output contains the estimated probability of being in each DFactor level, the factor mean, and standard errors for these estimates.



For each CFactor and GCFactors, this file also contains the factor means and for GClasses the classification probabilities and the modal assignment.

Covariate Classification. Classification is usually performed based on all available information for a case (Standard Classification). However, it is also possible to compute the probability of being in a certain latent class (or a factor mean), given covariate values only. In fact these are model probabilities; that is, $P(x|z)$ (see Technical Guide). These probabilities are useful for classifying new cases for which information on the dependent variable or indicators is not available.

Each row in the Covariate Classification output corresponds to a distinct pattern of active covariates that is observed in the data file. For Cluster and Regression models, each of these rows contains the estimated probability of being in each class. For DFactor models, each row in the Covariate Classification output contains the estimated probability of being in each DFactor level, the factor mean, and standard errors for these estimates. In multilevel models, it contains the GClass probabilities given group-level covariates.

Note: Inactive covariates do not influence the classification probabilities and hence have no effect on this output.

Standard Classification information as well as *Covariate Classification* can be viewed as Tabular output and/or can also be output to an external file. Selection of these from the Output Tab produces the Tabular output. Selecting *Standard Classification* and/or *Covariate Classification* from the 'Output to File' section of the ClassPred Tab produces the external files which contain the classification information as new variables appended to a copy of the input file used for estimation. See below:

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

Estimated Values. Shows/hides Estimated Values in Output Window (Regression only) This output section contains the class-specific and overall predicted values for each predictor pattern.

Set Profile (Available only for Choice Models) See Latent GOLD Choice Manual). Shows/hides Set Profile in Output Window

Set ProbMeans (Available only for Choice Models) See Latent GOLD Choice Manual). Shows/hides Set ProbMeans in Output Window

Importance (Available only for Choice Models) See Latent GOLD Choice Manual). Shows/hides Importance in Output Window

Iteration Detail. Shows/hides Iteration Detail in Output Window. If this output is not selected, it still will appear if any problems are encountered during model estimation.

STANDARD ERRORS AND WALD

Choose one of four options:

The first three options specify different types of information matrices to be used in the computation of standard errors and Wald statistics. The fourth option suppresses such computations

Standard (Hessian)

The Standard method makes use of the second-order derivatives of the log-likelihood function called the Hessian matrix. This is the default option.

Robust (Sandwich)

The Robust method "sandwiches" the inverse of the outer-product matrix by the Hessian information. Standard errors and Wald statistics obtained by the Robust method are less affected by distributional assumptions about the indicators and the dependent variable.

Fast (Outer Product)

The Fast method approximates the information matrix using the outer-product of the first-order derivatives of the log-likelihood function. The Fast method may be used in models in which the other two methods are computationally intensive. In such cases, one can also suppress the computation of standard errors and Wald statistics.

None

This option suppresses the computation of standard errors and Wald statistics (option None). This option may be useful when estimating models containing an extremely large number of parameters, in which case computation of the second-order derivatives (used in Newton-Raphson, standard error computations and Wald statistics) may take a lot of time. By setting the *Newton-Raphson Iteration Limit* to 0 and setting *Standard Errors and Wald* to *off*, the estimation process for such large models is much quicker.

PREDICTION TYPE (REGRESSION ONLY)

In regression, the program reports Prediction Statistics. It is also possible to write predicted values to a file. Predicted values can be computed in three different ways.

Posterior Mean

Posterior Mean predicted values are defined as weighed averages of the class-specific predicted values using an individual's posterior membership probabilities as weights.

HB-like

As in Hierarchical Bayes, the HB-like predicted values are based on the Individual Coefficients, which are weighted averages of the class-specific regression coefficients with the posterior membership probabilities as weights.

Marginal Mean

Marginal Mean uses the prior membership probabilities as weights, which means that the observed values on the dependent variable are not used to generate the predictions.

Posterior Mean and HB-like prediction yield similar results. These methods give a good indication of within-sample prediction performance.

Marginal Mean prediction yields much lower R-sq values, but gives a better indication of out-of-sample prediction performance.

CODING NOMINAL

Effect (default). By default, the Parameter Output contains effect coding for nominal indicators, dependent variable, active covariates and the latent classes (clusters). Use this option to change to dummy coding.

Dummy Last. Selection of this option causes dummy coding to be used with the last category serving as the reference category.

Dummy First. Selection of this option causes dummy coding to be used with the first category serving as the reference category.

VARIANCE/COVARIANCE MATRIX

When the input data file is either an ASCII text file or an SPSS .sav file, this option outputs the variance-covariance matrix of the parameter estimates to an external file.

Output Filename. Upon selection of this option, a default filename appears in the box directly below the check-box. Use the browse button to change the filename and/or its save location. The format of the output file will be the same as that of the input file (ASCII or .sav).

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

The body of the output file contains the variances and covariances of all model parameters. Each row in this output file corresponds to a parameter. The first variable (column) on this file is a string variable 'Location' which provides a unique name for the parameter such as 'r0001c01'. The 3 right-most columns in the output file are variables called 'se', 'param' and 'Label', serve to define the parameters. For example, for this parameter (row), the string variable called Label might contain a label such as 'purpose : 1 || Cluster1' which means that this is the parameter estimate for with the 1st category of the indicator PURPOSE, associated with cluster #1. The variables 'param' and 'se' correspond to the estimate and standard error for this parameter as reported in the Parameters Output.

The remaining variables on the file reproduce the parameter names provided in Location and contain the variance/covariance matrix. For example, the entry in row 1 (Location = 'r0001c01') and column 'r0001c01' is the variance of this parameter estimate. The entry in row 1 (Location = 'r0001c01') and column 'r0001c02' is the covariance associated with parameter estimates 'r0001c01' and 'r0001c02'.

Note: Most users will not need to use this option. These quantities are useful in computing the standard error of a particular function of the parameter estimates. For further details see Section 7.9 of Technical Guide.

DEFAULT OPTIONS

Click **Default** to restore the Output options to their original program default values.

Click **Save as Default** to save the current output settings as the new default values.

Click **Restore to Defaults** to restore the Output options to their last saved default settings.

Click **Cancel Changes** to cancel any changes that have been made to the Output options and not saved.

For further details and examples, see section 2.4 of Technical Guide.

ClassPred Tab: Classification and Prediction Output to a file

CLASSIFICATION OUTPUT TO A FILE

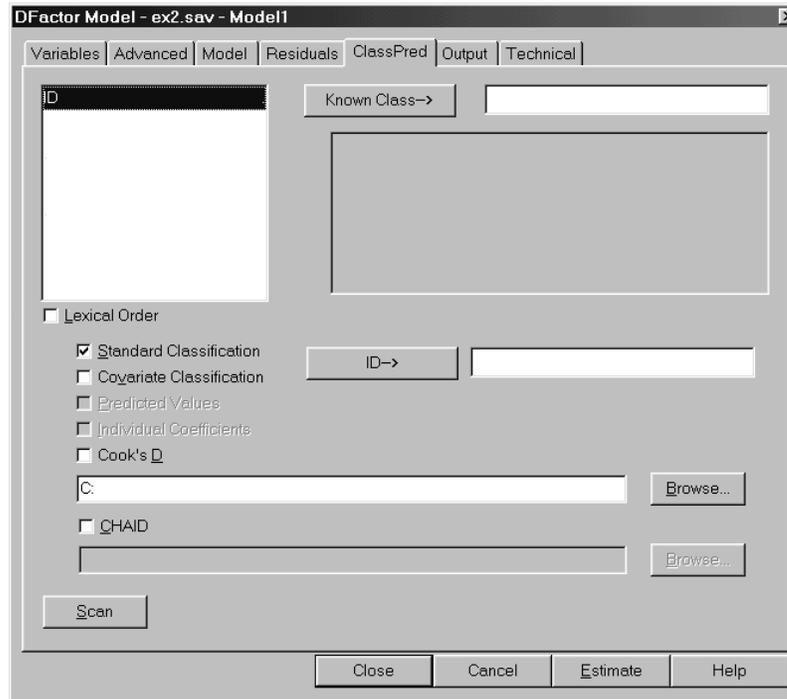


Figure 5-17. ClassPred Tab for DFactor Model

Standard Classification (optional)

When the input data file is either an ASCII text file or an SPSS .sav file, this option produces a new data file containing the standard classification information such as the probability of being in each cluster/class, or for DFactor models, in each level of each DFactor, together with any covariates and other variables specified in the Variables Tab and the variable included in the ID box of the ClassPred Tab (if any). The format of the output file will be the same as that of the input file (ASCII or .sav). This option is not available when using the range option in the Variables Tab to specify a range of models. In addition to these probabilities Latent GOLD also appends classification variables containing that cluster/ class/level into which the respondent should be classified (the one being the highest membership probability). For each case in the analysis file, the variables on the new data file consist of the model variables, the cluster/class/DFactor level probabilities, and class/DFactor classification (i.e., the index of the class/level containing the highest estimated probability). As an option, an ID variable can also be appended to the new file. For an example using this option, see Tutorial #3.

Output Filename. A default filename will appear in this box. Use the browse button to change the filename and/or its save location.

ID Variable. A single additional variable may be selected for inclusion (typically an ID variable or other key variable which provides a unique identification of each case on the file) to allow additional variables on the original data file that were not included in the analysis to be merged onto this file.

Note: The new file is created after the model has been estimated. After selecting this option, click Estimate to estimate the model and create the new file.

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

WARNING! This setting is not preserved across models - it must be selected explicitly for each model estimated.



For each CFactor and GCFactor, the corresponding factor means are output; for GClasses, the classification probabilities and the modal assignment are output.

The order of the variables in the output file (and labels for a .sav formatted file) are as follows:

- the Known Class Indicator (if specified in the ClassPred Tab)
- any covariates included in the model
- other model variables specified in the Variables Tab
- the variable included in the ID box of the ClassPred Tab (optional)

For LC Cluster models:

- clu#1 (Cluster1) - posterior membership probability for cluster 1
- clu#2 (Cluster2) - posterior membership probability for cluster 2
- (repeated for clusters 3,4, ...)
- clu# (Cluster modal) - the cluster number for the most likely (modal) cluster

For DFactor models:

- fac1#1: The probability of being in level 1 of factor 1
- fac1#2: The probability of being in level 2 of factor 1
- ... (repeated for levels 3,4, ...)
- fac1#: The level of factor 1 into which the case is classified
- fac1scr: The mean score for factor 1
- fac2#1: The probability of being in level 1 of factor 2
- fac2#2: The probability of being in level 2 of factor 2
- ... (repeated for levels 3,4, ...)
- fac2#: The level of factor 2 into which the case is classified
- fac2scr: The mean score for factor 2
- ... (sequence repeated for DFactors 3,4, ...)

- jfac#1 ('Joint DFactor 1 1')
- jfac#2 ('Joint DFactor 1 2')
- (repeated for joint levels '1 3', ... , '2 1', ...)

For LC Segmentation/Regression models:

- clu#1 (Class1) - posterior membership probability for class 1
- clu#2 (Class2) - posterior membership probability for class 2
- (repeated for classes 3,4, ...)
- clu# (Class modal) - the class number for the most likely (modal) class



Advanced:

- cfactor1 ('CFactor1')
- cfactor2 ('CFactor2')
- cfactor3 ('CFactor3')
- gclass1 ('GClass1')
- gclass2 ('GClass2')
- ...
- gcfactor1 ('GCFactor1')
- ...

Covariate Classification

Classification based on covariates, as is the case with Standard Classification information can be output to an external file. Selecting Covariate Classification from the 'Output to File' section of the ClassPredTab produces the external files. The external file corresponding to the Covariate Classification information contains the new variables appended to a copy of the input file used for estimation.



For multilevel models, this output file also contains the GClass probabilities given group-level covariates.

PREDICTION OUTPUT TO A FILE

Predicted Values (Regression only)

Predicted values for the dependent variable can be output to an external file. The method used to determine the predicted values (pred_dep in output file) depends on the Predicted Values setting on the Output Tab (Posterior Mean, HB-like, or Marginal Mean). The predicted value is a mean, except for nominal dependent variables for which it is a mode. For categorical dependent variables, the full estimated probability distribution for each replication is reported.

Individual Coefficients (Regression only)

It is also possible to output posterior mean estimates for the Individual Coefficients to an external file. These are weighted averages of the class-specific effects, where the posterior membership probabilities of a case serve as weights. In the output file, the coefficients appear in the same order as in the Parameters Output and are labeled as b1, b2, b3, etc. Both the individual estimates (est_b1, est_b2, etc) and the individual standard deviations (std_b1, std_b2, etc) are provided.

Cook's D

The Cook's Distance measure may be output to an external file. This measure is used to identify cases that have a large influence on the parameter estimates. A recommended cut-off point for Cook's distance is four times the number of parameters divided by the number of observations.

CHAID (Requires a license to SI-CHAID 4.0)

This option creates a CHAID settings file (.chd file) from your model that can be then opened via the SI-CHAID 4.0 program. With this option, a CHAID (CHI-squared Automatic Interaction Detector) analysis may be performed following the estimation of any LC model in Latent GOLD 4.0. By selecting 'CHAID' as one of the output options, a CHAID input file will be constructed upon completion of the model estimation, which can then be used as input to SI-CHAID 4.0. For more information regarding CHAID, see section 1.6 in Chapter 1, and Tutorial #4 in Chapter 7.



For additional information on the output to file options, see Section 7.9 of Technical Guide.

Step 10: Estimate the Model, View Output and Continue



Once you have specified your model, to estimate the model you may:

- ▷ Click Estimate (located at the bottom of the Analysis dialog box)

If your analysis window was originally opened from a new model (as opposed to a previously estimated model) you may estimate the model in other ways as well:

- ▷ From the Model Menu, select Estimate,

or

- ▷ Click 

If you opened a previously saved .lgf file and the Outline Pane contains names for 1 or more models that have not yet been estimated, to estimate all models that have not yet been estimated.

- ▷ click on the data file name in the Outline Pane
- ▷ select 'Estimate All' from the Model Menu

STOPPING MODEL ESTIMATION (MODEL MENU)

Upon beginning a model estimation, the stop button on the toolbar becomes red (this may take several seconds or longer) which indicates that it is now active. You can stop a model estimation once it has begun to accomplish either 1) canceling the estimation, or 2) pausing the estimation to view preliminary results and/or make changes to the requested output options or change the iteration or convergence limits prior to resuming estimation.

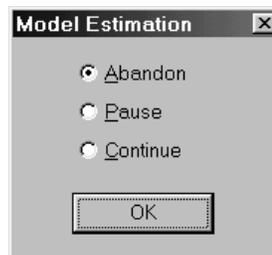


Once the stop button becomes active,

▷ To stop the estimation procedure select Stop from the Model Menu
or

▷ click on the  button in the toolbar

and a popup-menu appears:

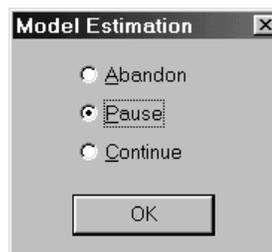


Abandon/Cancel Estimation of a Model

▷ Select **Abandon** to cancel the model estimation

This option returns the program to its state prior to beginning the Estimation. Output is produced only for models that completed the estimation process without being terminated.

Pause Model Estimation



The Pause option allows you to pause a model after the model estimation process has begun but prior to the estimation being completed to review preliminary Model Summary Output as well as any of the following Model

Output Sections that were requested in the Output Tab:

Parameters

Profile

ProbMeans

Bivariate Residuals

Iteration Detail

The Pause option is not available during estimation involving a bootstrap procedure.

If you are estimating a range of models, the pause option will pause the estimation process for the model currently being estimated and will cancel any further models. Depending upon how large the model is that is being estimated, it may take anywhere from one second to several minutes (or longer) to generate the preliminary output listings.

The Pause option does not cause a preliminary version of output to a data file to be created even if such was requested using the ClassPred Tab. If such has been requested, it will be produced only if the estimation is resumed and allowed to complete. See Resuming Estimation of a Paused Model below.

Prior to resuming model estimation, you may modify the Output options that were requested in the Output Tab or change the Iteration or Convergence Limits in the Technical Tab.

After a model is paused, the model name associated with that model appears in the Outline Pane with the characters 'Paused' appended to it. To change the Output options or the Iteration or Convergence limits, double click the model name to open the analysis dialog box and make the desired changes in the Output and/or Technical Tab. Note that the label 'Resume' replaces 'Estimate' on the Estimate button.

Continue Model Estimation

If Stop was selected in error, click Continue to continue estimating the model.

Resuming Estimation of a Paused Model (Model Menu)



After viewing the preliminary output and making changes to the options as described above, to Resume the estimation of a paused model:

▷ select Resume from the Model Menu

or

▷ Click the Resume button at the bottom of the Analysis Dialog Box associated with the Paused model.

▷ To open the analysis dialog box for a Paused model double click the name of the paused model. In the Analysis Dialog box for a Paused model, the word 'Resume' replaces the word 'Estimate' on the Estimate button.

or, you may

- ▷ Click on the name of the Paused Model and select Resume from the Model Menu.

BOOTSTRAP P-VALUE (L²) (MODEL MENU)

This option is only available if chi-squared statistics are available for the estimated model. Thus, it is not available:

- For any multilevel model
- For Cluster and DFactor models, if one or more indicators is specified as continuous.
- For Regression, if the dependent variable is specified as continuous.

With sparse data, the chi-squared based estimation for the p-value associated with L² cannot be trusted because these statistics do not follow a chi-squared distribution. When chi-squared statistics are available, a good alternative is to estimate the p-value by bootstrapping, or Monte Carlo simulation. The *Bootstrap L²* procedure involves generating a certain number of replication samples from the maximum likelihood solution and re-estimating the model with each replication sample. The bootstrap p-value is the proportion of replication samples with a higher L² than in the original sample.

A bootstrap p-value for the L² statistic can be obtained for a model after that model is estimated.



Once you have examined the results for an estimated model and wish to obtain a bootstrap p-value for it,

- ▷ right-click on the selected model to retrieve the model selection menu.

If chi-squared statistics are available for this model, the bootstrap option will be active on this menu.

- ▷ Select **Bootstrap L²**

The model will now be re-estimated using only a single start set and the best start seed, which guarantees that the bootstrap procedure starts with the previously estimated solution. Following the re-estimation of the model in the usual way, the estimation procedure for the bootstrap p-value begins. The number of Replications and the Seed used will be the ones provided in the Bootstrap section of the Technical Tab (see Bootstrap Options in Step 8).

After each replication of the procedure has completed, the bootstrap p-value is updated in the status bar displayed in the lower left portion of the screen.

Once the bootstrap procedure has completed, the name of the associated model together with the appended characters 'Boot' appears at the bottom of the Outline Tab. For example, if the estimated model is named 'Model3', the resulting model containing the bootstrap p-value will be named 'Model3Boot'. The resulting bootstrap p-value and standard error appear in the Model Summary Output in the columns to the right of the tradi-

tional p-value associated with this model name. All other output associated with the Boot model name is identical to that obtained earlier for the estimated model.

Note: Bootstrap information is provided only in the Model Summary Output listing. This information may be saved by copying it to the clipboard and pasting it elsewhere, or by using the Save Results option in the File Menu. If the Save Definition option in the File Menu is used to save a .lgf definition file for this model, the Bootstrap Seed is saved. Upon re-estimating the model following a File Open of the saved .lgf file, the model will again be estimated in the usual way but the bootstrap will not be performed. To reproduce the bootstrap, you will need to select Bootstrap L^2 from the Model Menu for the re-estimated model. Use of the saved Seed guarantees that the same replication data sets will be generated and you will be able to reproduce the bootstrap results obtained previously.

PROBChi Calculator (View Menu)

For assistance in assessing the improvement of a new model that imposes one or more testable restrictions on a previously estimated model, the ProbChi option allows you to obtain a p-value for a given chi-square (or vice-versa) for a specified number of degrees of freedom, df. For two models, Model2 being a more parsimonious (restricted) form of Model1, this tool tests whether the simplification is warranted using the difference between the L^2 statistics for the two models, $L^2(\text{Model2}) - L^2(\text{Model1})$, with $df = df(\text{Model2}) - df(\text{Model1})$. Enter the L^2 difference and the corresponding df and then click Chi->p to see if the difference between the models is significant. If not, you may use the more parsimonious model (Model2) without a significant loss of information.



To use this tool:

- ▷ From the menus choose View → ProbChi
- ▷ Enter a Chi-square value (or p-value) and number of degrees of freedom.
- ▷ Click on Chi->p (or p->Chi).

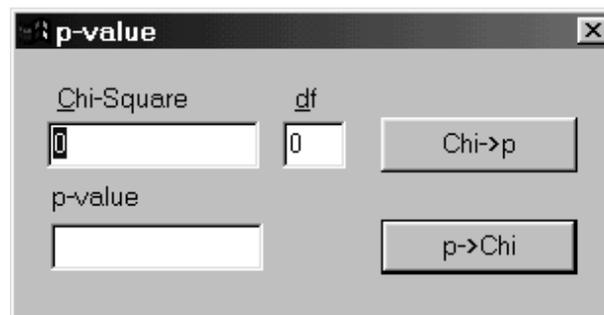


Figure 5-18: ProbChi dialog box

Given that the less restricted model is true, if the difference in model fit between the 2 models is not significant, the more restricted model may also be accepted as true and therefore preferred on grounds of being more parsimonious. On the other hand, if the difference is significant, the restricted model would be rejected in favor of the unrestricted model.

Note: To be valid, the chi-square value entered should only be an L^2 value or a difference between two L^2 values (as described above).

WARNING: The L-square difference is not a valid way to test the significance of adding 1 or more classes, DFactors, or levels associated with one or more DFactors to a model. To test these requires the more general conditional bootstrap approach.

CONDITIONAL BOOTSTRAP (BOOTSTRAP -2LL DIFF) (MODEL MENU)

This option requires that you have already estimated at least 2 models, one of which is selected for the conditional bootstrap calculation. The second model, selected from a list of eligible models, must be a restricted version of (i.e., nested within) the selected model.

Note: This option is not available for multilevel models.

Similar to the use of the PROBChi calculator, the conditional bootstrap option (abbreviated as 'Bootstrap-2LL Diff') provides assistance in assessing the improvement of a model that imposes one or more testable restrictions on another estimated model by estimating a p-value. However, it is much more general than the -2LL difference test, and can be used to formally assess the statistical significance of imposing any set of model restrictions. Given that the less restricted model is true, if the difference in model fit between the 2 models is not significant, the more restricted model may also be accepted as true and therefore preferred on grounds of being more parsimonious. On the other hand, if the difference is significant, the more restricted model would be rejected in favor of the less restricted model.

A common application would be to test whether an additional class, DFactor or DFactor level provides a significant improvement in model fit. For example, since a 3 class model can be obtained from a 4-class model by restricting the size of the 4th class to be zero, the conditional bootstrap option can be used to test whether there is a significant difference in model fit between these 2 models. If the difference is not significant, the simpler 3-class model may be used in place of the 4-class model. If the difference is significant, you could conclude that the addition of the 4th class provides a significant improvement.

Another important application is in the testing of order restrictions, such as the ordered-restricted clusters (Cluster) and order-restricted predictor effects (Regression) assumptions. The test would then be between a model with and a model without the ordered restrictions of interest.



To use this option to test the significance of the difference in fit between 2 model:

- ▷ Estimate 2 or more models, at least one being a simpler, restricted form of another.
- ▷ After completion of the estimation, right click on the less restricted model.

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

The available Model menu options appear.

- ▷ Select 'Boot -2LL Diff'

A list containing a subset of the other estimated models appears, containing possibly one or more simpler (nested) models that may be eligible to be used as reference models.

- ▷ Select the desired less restricted model

- ▷ Click OK

These steps are illustrated in Tutorial #1 in chapter 7 to test whether the improvement from a 3 to a 4-class model is statistically significant.

Once the conditional bootstrap procedure has completed, the names of 2 models -- the less restricted model and the reference model together with the appended characters 'Boot' -- appears at the bottom of the Outline Tab. The resulting conditional bootstrap p-value and associated standard error appears in the Model Summary Output in the columns to the right of the traditional p-value associated with the less restricted (selected) model. In addition, if chi-squared statistics are available, the bootstrap L^2 p-value and standard error appear in the Model Summary Output in the columns to the right of the traditional p-value associated with the eligible reference model. All other output listings associated with these 2 'Boot' models is identical to that obtained earlier for the 2 models estimated originally.

Note: In order for the results of the conditional bootstrap to be valid, the reference model must be a restricted form of (i.e., nested within) the original model. You should be aware that the list of eligible models may well include some models that do not qualify as nested, and hence should not be selected. For example, you should make sure that both models include the same variables (dependent, exposure, caseid, indicators, covariates, predictors, attributes, case weights, replication weights, known class, etc.).

Note 2: Bootstrap information is provided only in the Model Summary Output listings. This information may be saved by copying it to the clipboard and pasting it elsewhere, or by using the Save Results option in the File Menu. If the Save Definition option in the File Menu is used to save a .lgf definition file for both the source model (say named 'Model2Boot') and the selected reference model (say 'Model1Boot'), the Bootstrap Seed is saved (for both models). To reproduce the bootstrap results, upon re-estimating both models following a File Open of the associated saved .lgf file(s), you will need to select 'Bootstrap -2LL Diff' from the Model Menu for the re-estimated source model and select the appropriate reference model again as described above. Use of the saved Seed guarantees that the same replication data sets will be generated and you will be able to reproduce the bootstrap results obtained previously. The 2 additional models estimated will have 'Boot' again appended to the original names, so the reproduced bootstrap models will be named for example 'Model1BootBoot' and 'Model2BootBoot'.

OUTPUT FILES CREATED

Once a model has been Estimated or Paused, a number of output file listings are created which may be viewed, copied to the clipboard, or printed. Each file may be selected by highlighting the file name listed in the Outline pane. See Chapter 6 for specific details about the output produced. Following review of the output, you may elect

to estimate a new model.

DEFINING A NEW MODEL

Once you have estimated at least one model, there are several ways to specify a new model on the same data file.

After estimating a model, Latent GOLD automatically creates a new model name in the Outline pane (it will be the last model name listed for a data file).



To setup a new model,

- ▷ Right click on the new Model<n> name in the Outline pane to open the Model Selection menu (which will have a checkmark next to the last type of model estimated) and proceed from Step 2 above.

When this method is used, the default new model settings will be the same as the last estimated model of that type, with the following exceptions:

- The Seed in the Start Values section of the Technical Tab is restored to '0' (random seed)
- The Random Sets value in the Technical Tab is restored to the original default
- Any output-to-file selections specified in the ClassPred Tab for the estimated model are removed
- ▷ Double click on the model name for a previously estimated model. The analysis dialog box will open and contain the settings for that model. Change whatever settings you wish to change, then click Estimate. The new model will appear at the bottom of the model list for that data file.
- ▷ Choose Define from the Model Menu after highlighting a previously estimated model. The analysis dialog box will open and contain the settings for that model. Change whatever settings you wish to change, then click Estimate. The new model will appear at the bottom of the model list for that data file.
- ▷ Highlight any model name and select the type of model from the Model Menu. A checkmark specifies either the type of model last selected (for new models) or the type of model selected for a particular model name (for previously estimated models).

After estimating several models, you may wish to delete some of these models, or rename models to be more informative names. The following model management options are available:

CHANGING MODEL NAMES



To customize the model name for your model:

CHAPTER 5. BASIC STEPS FOR MODEL DEVELOPMENT

- ▷ Single click the model name (for example, 'Model1') in the Outline pane to highlight it.
- ▷ Single click on it again to edit the name (the name will be in an edit box).

Model names can be edited at this time before or after the model is estimated

Once you have an edit box open, you may also right click to open a menu that allows you to Cut, Copy, Paste, and Delete the model name. This menu also allows you to Undo your last edit or select the entire name for editing (Select All).

DELETING MODELS



To delete a previously estimated model:

- ▷ Single click the model name in the Outline pane to highlight it.
- ▷ Choose Delete from the Model Menu.

The model will be deleted from the Contents pane, and will no longer appear in the summary output when the corresponding data file is highlighted.



To delete all models associated with a data set:

- ▷ Single click the data file name in the Outline pane to highlight it.
- ▷ Choose File Close from the Model Menu.