First-Order Derivative Warning Message, Condition Number, and Non-Identification

Tihomir Asparouhov & Bengt Muthén

August 8, 2022

This note discusses the following warning message that sometimes appears in the Mplus output

THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE FIRST-ORDER DERIVATIVE PRODUCT MATRIX. THIS MAY BE DUE TO THE STARTING VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION. THE CONDITION NUMBER IS 0.889D-10. PROBLEM INVOLVING THE FOLLOWING PARAMETER:

A technical background is given, followed by a description of several investigation strategies. We then provide an extensive list of common examples where the message appears and provide guidance on how to resolve the specific issue.

1 Technical Background

To compute the maximum likelihood estimates, the log-likelihood function L is maximized with respect to all model parameters. If proper maximization has been achieved, the negative of the matrix of the second derivatives, also known as the Fisher information matrix, $-L''$ should be a positive definite matrix. The information matrix is inverted $(-L'')^{-1}$ to obtain the standard errors. If the model is not identified, the information matrix is singular, i.e., it is not invertible and the determinant is 0. Mplus performs a check to determine if the information matrix is invertible, i.e., it is not singular. In principle, to do that one can simply compute the determinant of the matrix and check that it is 0. Numerically, however, the matrix will not be precisely zero. Mplus uses 15 decimal digits of precision for every number. In addition, round-off error can accumulate during the computation and thus it is not uncommon for a number that is theoretically zero to become as large as 10^{-10}. In addition, computing the determinant of the information matrix to check the singularity of the matrix, is not a good idea as it exposes the computation to the scales of the observed variables. Instead of using the determinant to check for singularity, Mplus uses the condition number of the matrix. This is defined as the ratio of the smallest eigenvalue to the largest eigenvalue. If the matrix is singular the condition number is 0. Mplus uses as a cutoff value 10^{-10}, i.e., if the condition number of the matrix is less than 10^{-10} Mplus concludes that the matrix is close to being singular which indicates that the model might not be identified. The cutoff value can be changed using the CONDITION option in the ANALYSIS command. The condition number can be found in the Mplus output for every model. Small condition numbers should be viewed as a problem that may or may not need to be addressed. The smaller the condition number, the flatter the likelihood of the model is, i.e., the weaker the identifiability of the model, i.e., the data set contains little or no information for some of the model parameters. The smaller the condition number, the bigger the round off error in the matrix inversion, and as a result the bigger the round off error in the standard errors.

In addition to using the information matrix to check model identifiability, Mplus also uses the first-order derivative product matrix $\overline{L'(L')^T}$ which provides an approximation to the information matrix and is available in Mplus as the MLF estimator. This is the product matrix referred to in the above error message. Experience has shown that the MLF method is the best method to catch an unidentified model. It is the most accurate in terms of false negatives, i.e., least likely to not catch an unidentified model. Thus, Mplus performs the MLF singularity check even when the estimator that is used for the estimation is not the MLF estimator but is the MLR or the ML estimator. Unfortunately the MLF estimator also has a larger false positive error than the ML and MLR estimators. That is, the MLF check may report potential non-identifiability in situations when the model is identified.

When the MLF check fails, Mplus will produce an error message described above. The parameter number that is listed may be the parameter that is not identified or may be closely connected to that parameter. The unidentified parameter can also be a parameter that is close to the listed parameter in terms of its TECH1 order. The way Mplus identifies the problematic parameter is by sequentially analyzing the information matrix. First we analyze the first parameter only, then the first and the second, then the first three, etc. We conclude that the problematic parameter is the one that when added, the condition number drops below the cutoff value. Sometimes that drop will not occur at the exact unidentified parameter but at the next one (or the next few) as the condition number continues to decrease as we add more and more parameters.

When the above error message appears, Mplus can not guarantee that the model is identified. The identifiability of the model must be verified separately. For example, if the model is a well known model that has been used successfully with other data sets, the error message can be ignored.

In most situations, a condition number below 10^{-12} indicates a true non-identification. A condition number between 10^{-10} and 10^{-12} is usually an indication of some other problem that could be addressed by model or data modification.

2 Investigation Strategies

In this section we describe 4 general investigation strategies that may be helpful in identifying the source of the problem.

- 1. Using the MLF estimator
  The first strategy is to rerun the analysis using the following estimation settings: ANALYSIS: ESTIMATOR=MLF; CONDITION=0;. Using this strategy, some of the parameters will get huge / large standard errors. These are usually reported as actual large numbers or they are reported as ******. These parameters are directly responsible for the problem. With this investigation strategy we narrow down the identifiability issue to just one or several parameters rather than the entire model. All or some of these parameter could be fixed to immediately resolve the problem.

  There are three different scenarios that can easily be recognized when using this approach. The first scenario is when a parameter p converges to $\pm\infty$. Precise value for the parameter can not be established. The log-likelihood is virtually indistinguishable when the parameter takes any large value. Often such parameters are automatically fixed by Mplus, but in some situations they are not. Such parameters can be manually fixed to large values or the issue can be ignored, assuming that proper interpretation is placed regarding what the value is and why the standard error for this parameter does not carry any information. The second scenario is when the model implies a deterministic relationship between several parameters. For example, due to the way the model is set up, parameters p_1 and p_2 may be tied to a deterministic relationship such as $0 = p_1 + p_2$, i.e., the relationship will hold regardless of what data is used. In this case, both parameters will get large standard errors, but the exact deterministic relationship will not be revealed. If it is possible to determine the deterministic relationship between the parameters, the model can be augmented by a model constraint that accommodates the deterministic relationship. In the example, the model constraint will contain exactly $0 = p_1 + p_2$. This way the model will contain one less parameter and the singularity will be resolved. The third scenario is the case where parameters can not be identified but only a particular part of the parameters can. For example, the likelihood may depend on parameters p_1 and p_2 only through the difference of the two parameters, i.e., $b = p_2 - p_1$ is the identifiable quantity, but both p_1 and p_2 are not. The likelihood value obtained for p_1 and $p_2 = p_1 + b$ depends on b but not on p_1. In such a hypothetical example, both p_1 and p_2 will get large standard errors when we use the MLF estimator. If we can determine the precise relationship that holds and what exactly the identifiable quantities are we can resolve the issue. In the example simply fixing the parameter p_1 to any value will solve the identification problem. Here again we reduce the number of model parameters by 1 parameter without affecting the log-likelihood value and the model fit.

The technical background for strategy 1 is as follows. The CONDITION option refers to the cut-off value which determines if the information matrix is considered invertible. By default that setting is 10^{-10}. Setting it to zero means that the matrix will be inverted anyway even if it appears to be singular, due to having a low condition number. The MLF estimator has the neat property that the condition number is always positive, i.e., it is never negative. That means that numerically the matrix can always be inverted, even when the matrix is technically singular and the inversion essentially uses division by zero. The outcome of inverting a singular matrix like that is that some of the resulting values may be huge numbers. This property does not hold for the ML and MLR estimator where the condition number can be -1.0^{-15}. Even if we wanted to invert such a matrix it would be impossible because it would require taking a square root of a negative number.

- 2. Using the STARTS/OPTSEED
  The second investigatory strategy is to use a larger STARTS value so that the best solution is replicated at least twice. At that point, comparing the models using the top two OPTSEEDs will reveal if the point estimates are reliable. If the two models have the same point estimates, we can be confident that they are reliable. If the point estimates are different then clearly they are unreliable and there is a model identification problem. If the parameter estimates are different, most likely we have a version of the third scenario described above for the MLF strategy.

- 3. Simulation study
  A third strategy that can further illuminate the problem is to conduct a monte carlo simulation study with the model in question and a large sample size, possibly using the model parameters equal to the empirical estimates. Here we can estimate multiple replications to see how often such a problem occurs with simulated data. If 100% (or a high percentage value) of the replications yield similar identification problems we can conclude that indeed something is wrong with the model. If 0% of the replications have an identification problem, then we can conclude that the issue is with the data. It might be useful to then conduct simulation studies with small sample size or sample size similar to the empirical data to see if the problem starts to show up. In some cases, a small percentage of the replications result in identification problems. In that case, we can conclude that the identification problem applies to a particular set of parameters. We usually refer to this as empirical non-identification. Simulation studies closely resembling empirical analysis can also be used to verify that even when the MLF warnings occur, the empirical estimates are fine and the warning can be ignored. We can make this conclusion if the simulation study results show that the point estimates are unbiased and that the coverage is near the nominal 95% level. This happens in many situations. The MLF warning

is quite sensitive and can often produce false positive results, i.e., the MLF warning may indicate that there could be a problem but the ML or MLR point estimates and standard error are actually good.

- 4. Changing the estimator or estimation settings
  Mplus offers a variety of different estimators and estimation settings. Due to the use of different algorithms, a different estimation setting may result in a more revealing error message. In our experience the most successful of these are: adding ANALYSIS: ALGO=INT, or ANALYSIS: ESTIMATOR=BAYES.

3 Examples

Here are some common causes of the message and how they can be resolved.

3.1 Binary variables treated as continuous

To avoid listwise deletion due to missing data on covariates, the covariates are often brought into the model by mentioning their means, variances, or covariances. With categorical variables, this means that they are treated as continuous variables. With binary variables, the MLF message gets triggered in this case because the mean of the variable p is directly related to the variance $p(1-p)$ (one parameter p exists while two are estimated). In almost all situations this should be addressed by model modification. Treating the variable as continuous is not ideal. Missing data estimation for continuous variables assumes normal distribution and this misspecification will likely (depending on the amount of missing data) bias the estimation. There are two options in this case. Option one is to model the variable as categorical. Option two is to use multiple imputations where the variable is specified as categorical so that the imputed values are also binary (see User's Guide example 11.5 for how to impute missing values in Mplus).

Another situation where a binary covariate is treated as continuous is the situation where the variable is used in a WITH statement with all other covariates to ensure that correlations between the covariates are taken into account. However, if a variable is independent, it is automatically correlated with all other covariates in the model and such WITH statements are unnecessary and should be removed.

If a dependent binary variable is treated as continuous, causing the MLF message to appear, the variable should simply be declared as categorical.

3.2 Variables in the model are on very different scales

Sometimes variables have very different variances due to being measured on different scales and this may trigger the MLF message due to a low condition number. The scales of the variables can be changed so that they are more similar and that will improve the condition number. The scale of the variables can be changed either by DEFINE: Y=Y/10; or by DEFINE: STANDARDIZE Y.

3.3 There are more parameters in the model than observations or categorical patterns

In this case the MLF check is triggered 100% of the time. It is possible to verify that the model is identified with a bigger data set. One way to do that is to use a data set that consists of multiple copies of the original data set. If the model is a standard well known model the MLF warning should simply be ignored. If that is not the case, one can conduct a simulation study with a larger number of observations to verify that the model is identified. If the model is identified, the standard errors can be considered reliable.

If all observed data is categorical, the number of observations in this discussion is replaced by the number of different patterns. The number of different patterns can usually be obtained with OUTPUT:TECH10 using a simple model (if the current model doesn't produce that already). What this means is that for the purpose of identifiability, observations with the same outcome on all variables will not count as different. As an example, 5 binary variables can produce no more than 32 patterns. Any model with 32 or more parameters is a non-identified model, regardless of the actual data set. If a particular data set has 15 different patterns, i.e., 17 of the remaining possible patterns were not observed in the data, a model with 15 or more parameters will also produce a warning message and will be unidentified. While the model may be identified with a different data set, numerically, the model is not identified in this data set. When the number of parameters in the model is at least as big as the number of categorical patterns, the model is generally unidentified and at least a portion of the parameters are likely to have unreliable point estimates and large standard errors.

3.4 There are more parameters than clusters for type=complex, type=twolevel or type=threelevel

In this case the MLF check is triggered 100% of the time. It is possible to verify that the model is identified with a bigger data set. In three-level models, the number of parameters is compared against the highest level number of clusters. In multiple group analysis the number of clusters in each group must be at least as big as the number of parameters identified by that group to avoid the MLF warning.

Also note that even if the number of clusters is slightly more than the number of parameters, the MLF matrix may in some cases still become singular for various additional reasons, which can generally be summarized under the umbrella that some of the clusters did not provide enough information for the parameters. This is not an unusual occurrence. We do not recommend investigating which particular cluster did not provide additional information, this would not inform us of what to do next. The bottom line is that the number of clusters is too low, lower than the number of parameters, or barely above that number.

In principle, when the MLF message occurs and the source of the problem is precisely the low number of clusters, one may be able to ignore the message. In some cases, we can be certain that the model is identified simply because it is a well known model. This would then implicate the low number of clusters as the most likely source of the warning. If this is the case, the model results and standard error can actually be trusted. Simulation studies confirm that the standard errors obtained by ML and MLR estimators are fairly reliable, even when the MLF warning occurs due to low number of clusters. That said, however, the quality of the standard errors is directly tied to the number of clusters. If that number is low, asymptotic standard error can be somewhat imprecise. A useful action (although not required) that can be taken in such a situation is to simplify the model so that the number of parameters becomes much smaller than the number of clusters. For example, a group variable can be converted to a predictor.

3.5 Parameter appears to be converging to a large value

Mplus will often identify parameters that are causing identifiability problems, in terms of causing low condition numbers, and fix those parameters.

For example, when a binary indicator for a particular class is estimated to be a perfect indicator for that class, the threshold value will become +- large values. Such thresholds are equivalent to +- infinity and are essentially unidentified but are safe to fix. By fixing these parameters Mplus preserves the quality of the model and avoids the low condition number. Another example is the situation when a categorical variable is regressed on another categorical variable (latent or observed) and there are empty cells in the joint distribution of the two variables. Some of the regression parameters may be estimated to +- infinity and are safe to fix as well. Mplus automatically fixes threshold parameters if they reach the values of +-15 (this is controlled by the LOGHIGH and LOGLOW options of the ANALYSIS command). If a threshold is estimated at 14, it will not be fixed by Mplus and may produce a low condition number. You can change the LOGHIGH/LOGLOW options or ignore the MLF warning message. In this situation it is useful to look at and report the model parameters in the probability scale as those are well identified and do not have values converging to +- infinity.

3.6 Cluster invariant variable in two-level models

Suppose that a within-only variable in a two-level model takes the exact same values in each cluster. An example of such a variable is TIME in a two-level longitudinal model where the cluster is the person and the observations within the cluster are the observations at particular time points. If TIME takes the exact same values in each cluster and is modeled as a dependent variable, it will result in an MLF warning message due to a lack of variation in the score of TIME's parameters. The TIME variable could inadvertently be made into a dependent variable if it is correlated with other variables in the model. In this case the message can be ignored or the TIME variable can be treated as a covariate instead of a dependent variable. If a variable is correlated with the TIME variable, it should be regressed on it rather than correlated with it to preserve the exogeneity of the TIME variable.

3.7 Group-mean centered within-only variable with estimated intercept or mean

Suppose that a within-only variable Y in a two-level model has been group-mean centered using the command
DEFINE: CENTER Y(GROUPMEAN);

This command essentially removes the between part of the variable. The cluster sample mean is subtracted from each observed value in the cluster. The centered variable Y has a cluster sample mean of 0 in every cluster. This kind of centering is necessarily followed by a within-only specification
VARIABLE: WITHIN=Y;
(otherwise the between part will be estimated to the constant 0). The issue that arises here is that not only the random part of the intercept is removed, but also the fixed part of the intercept is removed. Therefore, the mean parameter of Y is necessarily 0. If that parameter is not fixed to 0, but is estimated as a free parameter, it is very likely that the MLF warning will appear. There are instances where it won't appear. For example, if the variable Y is regressed on another variable X that is not group-mean centered, the warning will not appear (even though the underlying modeling problem is still there and should be fixed).

The MLF warning should not be ignored in this situation . The proper model setup is to have the intercept of Y be fixed to zero and to also group-mean center all predictors of Y.

3.8 Class specific anomalies in Mixture models

In Mixture models, the classes are unknown and can shift in an unexpected way. When the estimation has completed, the standard errors are computed approximately as if the classes are determined and the computation is similar to known multiple groups. That means that any one of the above listed problems can occur for just one particular class, in addition to other problems. The most common situations that occur in Mixture models are as follows

- The number of observations in a particular class are smaller than the number of parameters estimated in that class. This is the same issue as in 2.3 but now it occurs in just one specific class.

- When a class specific regression is estimated such as Y on X, and one or both of the variables are categorical, the class formation may lead to the following unanticipated outcome: Var(Y)=0 or Var(X)=0. Both of these would imply that the class specific regression of Y on X is unidentified and the model must be adjusted accordingly. Using the option OUTPUT:TECH7 is very valuable in such cases as it will clearly identify this problem.

3.9 Parameters restricted by inequality constraints

Inequality constraints such as

$$F(v) > 0, \tag{1}$$

where v is the vector of model parameters and F is a function of these parameters can lead to the MLF warning message when the parameters have converged to the borderline solution, i.e., when $F(v) \approx 0$. This happens because Mplus implements the so-called slack parameter methodology. Instead of estimating the model under the inequality constraint, Mplus estimates the model

$$F(v) = Exp(a),$$

where a is a new model parameter. When the parameter estimates are near the border, the parameter a is near $-\infty$. At that point in the parameter space, small changes in a do not affect the likelihood and therefore the derivative is 0 and with a very high probability this will cause the MLF singularity warning. The slack parameter essentially creates the situation described in Section 3.5.

There is a simple resolution to the problem in this case. Instead of estimating the model under the inequality constraint (1), we can estimate the model under the constraint

$$F(v) = 0. \tag{2}$$

Because the original model estimation converged to the border anyway, the re-estimated model will have the same log-likelihood and will provide the same data fit. Note that when there are multiple constraints involved in the model estimation, only those that are at the border need to be converted to equalities. If there are multiple inequality constraints and the inequality constraints are complex, it may be unclear which inequality constraints are at the border. In this case, one can add $F(v)$ as a new parameter in MODEL CONSTRAINT. If that new parameter is near 0, clearly the parameter estimation has reached the border.

There is a secondary reason for converting the inequality constraint to an equality constraint when the parameters converge to that border. Asymptotic theory assumes that the likelihood is maximized at an interior point of the admissible space. When an inequality constraint is imposed in the estimation and the maximization has converged to the border of the admissible space, the asymptotic variance covariance is unreliable. Converting the inequality to equality resolves that problem. The standard errors obtained under the equality constraint are more reliable as they do not violate the asymptotic theory assumptions.

3.10 Empty middle categories in ordered categorical variables

This situation typically arises in Mixture models where in some classes a middle category of an ordered categorical variable does not occur in a particular class. In such situations two of the neighboring thresholds τ_j and τ_{j+1} become identical. The MLF matrix will in turn be singular as the asymptotic correlation between these two parameters will converge to 1. There are two possible resolutions. If the ordered categorical doesn't have any direct effects, i.e., the variable serves as a latent class indicator but there are no additional predictors for that variable, the most straightforward solution is to simply declare the variable as NOMINAL instead of as CATEGORICAL. If this solution is not possible, after identifying the identical thresholds by inspecting the output, this MODEL CONSTRAINT statement can be includes

$$\tau_{j+1} = \tau_j + 0.0001.$$

The model should be re-estimated with the above parameter constraint and then the asymptotic variable covariance matrix will not be singular. The pair of identical parameters has been reduced to just one parameter. The log-likelihood for the two models should be identical, possibly provided with starting values. Note also that parameter constraints must be used for every pair of identical thresholds.

3.11 Highly correlated variables

If deterministic relationships exist in the data, such as $Y_2 = Y_1 + 1$, where Y_1 and Y_2 are dependent variables in the model, the MLF matrix will be singular because the correlation between the means of the two variables will be 1. However, such pure relations are likely to cause even bigger problems, such as convergence problems, and are likely not going to reach the computation of the standard errors where the MLF error message appears. Deterministic relationships that hold for almost all observations, however, may converge but nevertheless have MLF singular matrix. For example, situations where $Y_2 = Y_1 + 1$ holds for all but one observation will have a non-singular sample variance covariance matrix which can facilitate model convergence. The MLF matrix, however, will still be singular because the number of non-deterministic observations is smaller than the number of parameters (specific to the addition of Y_2) that are estimated in the model. The resolution of this situation is to simply remove one of the variables that is highly correlated. Such variables contribute essentially no additional information and removing these from the analysis will cause no harm but will prevent identification problems.

3.12 General deterministic relations among the variables

This would essentially be just a generalization of the concept from the previous section but involving more variables. The most common scenario is if $Y_3 = Y_2 - Y_1$. In a case like this, the variable Y_3 should not be in the model in the first place. Modeling Y_1 and Y_2 is sufficient. The distribution of Y_3 should be derived from the estimated model, i.e., from the joint distribution of Y_1 and Y_2. If the variable Y_3 is kept in the model, the Y_3 distribution parameters will be dependent with a deterministic relationship on the model parameters of Y_1 and Y_2 and thus the variance covariance matrix will be singular.

If the data is prepared by the same person who analyzes the data, presumably the deterministic relationships among the variables will be known and can be avoided. In certain cases, the data is prepared by someone else, and the analyst inadvertently uses a model that can not be estimated due to such deterministic relationships. Here we describe a simple strategy that may help discover these relationships even when they are not known ahead of time. The first step is to locate the smallest model that contains the problem. In a series of model reduction we can eliminate all variables from the model that do not resolve the problem when removed. Suppose that in our example there are 10 variables $Y_1, ..., Y_{10}$. Removing one variable at a time, we re-estimate the model. If Y_1 is removed from the model the estimation problem will disappear. This leads to the conclusion that most likely Y_1 is involved in some kind of model violation. Removing Y_{10} from the model will result again in a convergence or a standard error problem. That means that Y_{10} can be permanently removed from the model for the purpose of identifying the model violation. The variable Y_1 can not be permanently removed from the model. Continuing with this process removing one variable at a time (or several variables at a time if the model has a large number of variables) we will arrive at the minimum number of variables that still exhibit the model violation. That set in our example will be a model for Y_1, Y_2 and Y_3. At this point there are several different options. One option is to manually inspect the data and look for simple deterministic relations such as one variable is the sum of several other variables, one variable is the product of two other variables, a sum of several variables is equal to a constant such as 0 or 1. Although this looks like a random strategy, in practice it usually works as most dependencies are quite simple. If this doesn't succeed we would recommend estimating this specific model Y2 on Y1; Y3 on Y1 Y2; ...., i.e., sequential regression where each variable is regressed on all prior variables (where the order of the variables is usually the order assigned in the USEVAR option). Estimating this model can quickly identify the deterministic relationship that is linear or close to it. In the results section for this model, we look for a residual variance that is 0 or near 0. Such variance indicates that there is a nearly perfect relationship among the variables and the actual model parameters reveal the relationship. In our example the residual variance of Y_3 will be zero, the regression coefficient of Y_3 on Y_2 will be 1 and the regression coefficient for Y_3 on Y_1 will be -1. This process will fully discover the relationship $Y_3 = Y_2 - Y_1$.