WU WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD
EQUIS ACCREDITED

ePub^WU Institutional Repository

Patrick Mair

A Framework to Interpret Nonstandard Log-Linear Models

Article (Published)
(Refereed)

http://epub.wu.ac.at/

A Framework to Interpret
Nonstandard Log-Linear Models

Patrick Mair
Vienna University of Economics and Business Administration

Abstract: The formulation of log-linear models within the framework of Generalized Linear Models offers new possibilities in modeling categorical data. The resulting models are not restricted to the analysis of contingency tables in terms of ordinary hierarchical interactions. Such models are considered as the family of nonstandard log-linear models. The problem that can arise is an ambiguous interpretation of parameters. In the current paper this problem is solved by looking at the effects coded in the design matrix and determining the numerical contribution of single effects. Based on these results, stepwise approaches are proposed in order to achieve parsimonious models. In addition, some testing strategies are presented to test such (eventually non-nested) models against each other. As a result, a whole interpretation framework is elaborated to examine nonstandard log-linear models in depth.

Zusammenfassung: Die Formulierung log-linearer Modelle im Kontext von Generalisierten Linearen Modellen führt zu neuen Möglichkeiten in der Modellierung kategorialer Daten. Diese Modelle sind nicht beschränkt auf die übliche Kontingenztafelanalyse mit hierarchischen Interaktionen und werden als nonstandard log-lineare Modelle bezeichnet. Bei solchen Modellen kann es vorkommen, dass die Parameter nicht mehr eindeutig interpretierbar sind. Dieses Problem wird im vorliegenden Artikel gelöst, indem die relativen Beiträge einzelner codierter Designvektoren ermittelt werden. Anhand dieser Ergebnisse werden schrittweise Verfahren präsentiert, um sparsame Modelle zu fitten. Zudem werden Ansätze gezeigt, um resultierende (evtl. nicht-genestete) Modelle gegeneinander zu testen. Somit wird eine Prozedur vorgeschlagen, die es erlaubt, solche nichtstandard log-linearen Modelle eingehend zu analysieren.

Keywords: Dominance Analysis, Stepwise Selection, Non-Nested Model Testing.

1 Introduction

Log-linear models (see e.g. Goodman, 1970; Bishop, Fienberg, and Holland, 1975) are used to model the frequencies of multidimensional contingency tables. With this modeling approach specific statements about the dependency structures among the variables (i.e. dimensions) can be made. The basic saturated formulation for a, say, three-way table is

$$\log(m_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} .$$

The expected cell frequencies are denoted by m with the cell index subscripts i, j, and k. λ is the intercept parameter (i.e. no effects). The main effects for variables A, B, and C

are λ_i^A, λ_j^B, and λ_k^C. The two-way interactions are λ_{ij}^{AB}, λ_{ik}^{AC}, and λ_{jk}^{BC}. The three-way interaction is given by λ_{ijk}^{ABC}. The common restrictions are that the main effect parameters sum up to 0. The same applies to the two-way interactions as well as for the three-way interactions for each index. Further explanations and issues of parameter interpretation in terms of the mean of the logarithms of the expected cell frequencies can be e.g. found in Bishop et al. (1975).

With the introduction of generalized linear models (GLMs) by Nelder and Wedderburn (1972) it was shown that log-linear models fit into the corresponding framework (see e.g. Evers and Namboodiri, 1979; McCullagh and Nelder, 1989) given by

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$$

with $g(\boldsymbol{\mu}) = \log(\mathbf{m})$ and $\mathbf{X}$ as the design matrix. As a consequence, this approach offers a more flexible tool in log-linear modeling. The corresponding log-linear models are so called *standard* and *nonstandard log-linear models* (see Rindskopf, 1990, 1999; Mair, 2006) and can be defined as follows:

Definition "Nonstandard log-linear models": If the design matrix $\mathbf{X}$ deviates from the ordinary effect or dummy coding scheme with corresponding main and interaction effects, $\log(\boldsymbol{m}) = \mathbf{X}\boldsymbol{\beta}$ can be considered as *nonstandard log-linear model*.

In other words, standard log-linear models are the direct "translation" of the Goodman representation into GLM. All other models are nonstandard. It is to mention that if some coding vectors are missing in order to fulfill the hierarchy principle (lower-order parameters can not be removed if higher-order parameters are still in the model; see e.g. Bishop et al., 1975) the model is non-hierarchical (see Magidson, Swan, and Berk, 1981). As suggested by the definition above, non-hierarchical models can be considered as special cases of nonstandard log-linear models.

A simple example of a nonstandard model is given in von Eye and Spiel (1996). The data describe repeat restaurants visitors' choices ($N = 94$) of main dishes at two occasions. These main dishes are Prime Rib (R), Sole (S), and Vegetarian Plate (V). The observed frequency vector of the underlying (3×3)-table is $\mathbf{n}^T = (25, 14, 3, 11, 17, 6, 7, 3, 8)$. In order to analyze a quasi-symmetry model for this table, they propose the design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 \end{pmatrix}$$

Obviously, the first vector is the intercept, followed by two main effect vectors for the first testing occasion and two main effects for the second occasion. So far, the model would be standard and the corresponding fit statistics would be $G^2 = 15.863$ ($df = 4$; $p = 0.0032$). Since there are three additional vectors for the symmetry pairs, the model

becomes nonstandard and the likelihood ratio (LR) value is $G^2 = 2.966$ ($df = 1$; $p = 0.085$).

For the variety of applications of such models see Mair and Eye (in preparation). In this paper a framework is presented that allows to analyze such nonstandard log-linear models in depth. This includes a stepwise approach to achieve a fitting model, determine the relative importance of single effects to model fit, and the comparison of two models (even on a non-nested level).

2 Establishing the Framework

The main problem in the analysis of nonstandard log-linear models is that, in general, the coding vectors in the design matrix are correlated among each other. By the way, this fact occurs in ordinary multifactorial ANOVA as well. A simple example should reflect this circumstance: Let us consider a (3×2)-table which leads to the saturated log-linear formulation with design matrix $\mathbf{X}$ (effect coding)

$$
\log \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \\ m_{31} \\ m_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}
$$

The first design vector in $\mathbf{X}$, here denoted by $\mathbf{x}_0$, corresponds to the intercept. The successive two vectors $\mathbf{x}_1$, $\mathbf{x}_2$ are the main effects for factor A. $\mathbf{x}_3$ is the main effect vector for B whereas $\mathbf{x}_4$ and $\mathbf{x}_5$ are the two-way interactions $[A : B]$. By examining the correlation structure of the effects it is obvious that $\mathbf{x}_1$ and $\mathbf{x}_2$ are not independent from each other. In fact, the correlation $r(\mathbf{x}_1, \mathbf{x}_2) = 0.5$. In contrast, it holds that $\mathbf{x}_1 \perp \mathbf{x}_3$ and $\mathbf{x}_1 \perp \mathbf{x}_4$. It is straightforward to expand this correlation concept also with respect to $\mathbf{x}_4$ and $\mathbf{x}_5$. However, there is a correlation structure in $\mathbf{X}$ that leads to the fact that the parameters are not interpretable independently from each other.

In the standard case, as above, this result is not grave. The parameters are still interpretable in terms of odds ratios (Rudas, 1988) or by deviations from the grand mean of the logarithm of the expected cell frequencies (Goodman, 1970). But in the nonstandard case the correlation structure becomes even more complex. The consequences are that the parameters are not interpretable independently from each other and furthermore, it is not warranted that they are meaningfully interpretable at all. As Mair and Eye (in preparation) show (see also Section 2.2), the formal representations can become exhaustive and thus it is difficult (if not impossible) to interpret these terms. In order to circumvent this ambiguous way to interpret the parameters, a numerical interpretation approach is presented in this paper. But first, a stepwise model build-up approach is discussed in the next subsection and, after determining the relative contribution of effects, the last subsection deals with the comparison of non-nested log-linear models. With these three issues integrated in a common framework programmed in R, standard as well as nonstandard models can be analyzed in depth.

2.1   Forward Selection of Coding Vectors

In ordinary stepwise regression analysis the selection criterion is based on a partial F-statistic. The predictor with the (significantly) largest gain in explained variability of the response Y is entered (forward stepwise selection). Due to multicollinearity issues an entered predictor can also be re-dropped. In log-linear modeling, i.e. dealing with categorical data, no sum of squares (SS) and, thus, no F-statistic can be defined. Several stepwise procedures have been elaborated (Brown, 1976; Benedetti and Brown, 1978; Edwards and Havranek, 1985). For the current purpose the forward selection presented by Goodman (1971) based on the χ^2-decomposition is adopted. That is because Goodman's approach does not require any hierarchy structure in the model effects entered, unlike the approaches mentioned above.

The starting points are base models that may derive either from certain preselections (see Brown, 1976) or by choosing a priori some non-saturated models. After choosing a base model $M^{(0)}$, those models $M_r^{(1)}$ ($r = 1, \ldots, R$) that differ in exactly one additional main or interaction effect from $M^{(0)}$ are considered. R is the number of these less parsimonious models. First, for each of the $M_r^{(1)}$-models the single LR-statistic is computed and for the best model $M_{r*}^{(1)}$, the conditional LR-statistic $G^2(M^{(0)} \mid M_{r*}^{(1)})$ is calculated. If either $G^2(M_{r*}^{(1)})$ or $G^2(M^{(0)} \mid M_{r*}^{(1)})$ are significant (or both), $M^{(0)} := M_{r*}^{(1)}$. Hence, $M_{r*}^{(1)}$ is the new base model and the procedure continues. Otherwise the algorithm stops at step $t = T$ and M^0 is the best model.

The only supposition for this method (Goodman, 1971) refers to the nested order of the models in order to calculate the conditional LR-statistic. This nested order is given for standard and nonstandard log-linear models since by operating stepwise and starting from a base model, one single parameter is added at each step. Therefore, the model at step $t - 1$ is nested in the model at step t ($t = 1, \ldots, T$). The final model is built up in the following way:

- Define an initial base model $M^{(0)}$. Usually, this model contains only the intercept parameter β_0, i.e. $\log(\mathbf{m}) = \beta_0 \mathbf{1}$, but, in general, any other model can act as $M^{(0)}$.

- Compute all conditional LR-statistics referring to each single effect. The effect with the highest gain in $G^2(M^{(0)} \mid M_r^{(1)})$ is added to $M^{(0)}$. The resulting model is $M_{r*}^{(1)}$.

- Set $M_{r*}^{(1)}$ as the new base model, i.e. $M^{(0)} := M_{r*}^{(1)}$ and add one additional parameter following the explanation in step 2.

- Proceed until neither $G^2(M^{(0)} \mid M_{r*}^{(1)})$ nor $G^2(M_{r*}^{(1)})$ is significant anymore.

A modification proposed here for a nonstandard model is the following: This part of the procedure acts as an extensive exploration of the effects. Thus, a researcher does not have to add the parameter with the highest gain in $(G^2(M^{(0)} \mid M_r^{(1)})$ by force. He can add a less attractive parameter (following e.g. certain hypotheses) and compare the resulting model with another model built up with this approach as well. In this sense, there also is not necessarily a need to proceed until non-significance of the LR-statistic is reached.

As mentioned, this stepwise approach does not require any hierarchical order in the effects following the *hierarchy principle* (see e.g. Bishop et al., 1975) which is usually considered in ordinary log-linear modeling. For instance, without entering the corresponding main effect one can enter an interaction effect. This is not consistent with the

mentioned hierarchy principle and would lead to a, so called, *non-hierarchical log-linear model* (see e.g. Magidson et al., 1981; Breen, 1984). The relevance of such a model from a content based view is not discussed here and, hence, just referred to Fienberg (1979) and Wickens (1989). In addition, it has to be remarked that from a GLM perspective such non-hierarchical models are nonstandard as well.

To conclude, the mentioned routine allows to enter arbitrary effects and to determine the goodness-of-fit after each step. In addition, a Pseudo-R^2 can be computed even though the interpretation of the value of R^2 is rather ambiguous for categorical data and should not be over-interpreted (Shtatland, Kleinman, and Cain, 2002).

2.2   Determining the Relative Contribution of Effects

As mentioned in the introduction, in general, the design vectors are correlated among each other. This implies that the parameters $\boldsymbol{\beta}$ are not interpretable independently from each other. Moreover, by imposing some nonstandard coding structure in the design matrix the question raises whether the parameters can be interpreted at all.

To clarify this issue, ordinary hierarchical models are regarded first. For these models, the discussions about the meaning of the interpretation of single parameters, unlike the interpretation of the whole selected model are manifold (see Holt, 1979; Wilson, 1979; Gilbert, 1981; and Alba, 1988). However, Rudas (1988) elaborates how the parameters can be interpreted in terms of odds ratios. For general tables (i.e. polytomous variables) with more than three dimensions the interpretation on the basis of odds ratios becomes very complex (i.e. conditional odds ratios of higher order). Also the ANOVA-like interpretation (see e.g. Bishop et al., 1975) as well as a related approach proposed by Elliott (1988) do not work straightforwardly for high-dimensional tables. Moreover, it has to be stressed that this discussion only refers to the hierarchical (standard) case.

In the nonstandard case, the interpretability becomes even more complex. Mair and Eye (in preparation) use the formal parameter representation derived from the GLM equation, i.e.

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\log(\mathbf{m}).$$

It is to mention that this statement (see e.g. Bock, 1975; Rindskopf, 1990) has nothing to do with *least squares estimation*; it is just a formal representation of the parameter vector. In fact, the parameters are usually estimated through *maximum likelihood* which can be performed with ordinary GLM compatible software like R, SAS, SPSS, and LEM (Vermunt, 1997). However, in nonstandard models the formal representation can become very complex and the parameters are not interpretable anymore on a formal and content based level, respectively.

Due to the mentioned difficulties in interpretation, the solution proposed in this paper is to focus on the effects in the design matrix and not on the parameter vector itself. The design matrix is established by the researcher and, therefore, the meaning of the design vectors is known. What is unknown is the relative importance and contribution of the single design vectors for model fit. In other words: How important is a certain design vector in the context of the other vectors in the model? Since the design vectors are correlated among each other, the following approach is proposed.

Table 1: Gender/Smoker Cross-Classification.

	Smoker	Non-smoker	Σ
Female	10	40	50
Male	20	30	50
Σ	30	70	100

To be able to deal with multicollinearity issues in ordinary regression, Budescu (1993) has elaborated a procedure for determining the mean importance of the predictors in the model (see also Azen and Budescu, 2003). This approach is called *dominance analysis* and is based on multiple correlations between a certain subset of predictors and the response. To be able to apply this concept to log-linear models the question has to be clarified whether it makes sense to compute a correlation between a design vector $\mathbf{x}_i$ (which is the predictor) and the expected frequency vector $\mathbf{m}$, respectively $\log(\mathbf{m})$ (which is the response). A simple hypothetical example can be considered to examine this issue. A cross-classification of gender and smoking habit leads to Table 1. The corresponding saturated log-linear model is

$$\log \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} .$$

$\mathbf{x}_0$ is the intercept, $\mathbf{x}_1$ the gender effect, $\mathbf{x}_2$ the smoker effect, and $\mathbf{x}_3$ the interaction. Since the model is saturated, $\mathbf{m}^T = \mathbf{n}^T = (10, 40, 20, 30)$. The correlations for the main effect vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are $r(\mathbf{m}, \mathbf{x}_1) = 0$ and $r(\mathbf{m}, \mathbf{x}_1) = -0.894$. From the marginal frequencies in 1 it is obvious that there is no gender effect; $r(\mathbf{m}, \mathbf{x}_1)$ reflects this topic. That there are more non-smokers than smokers is reflected by a negative $r(\mathbf{m}, \mathbf{x}_1)$, since the non-smokers are coded by -1. The same can be done for the interaction term and for any other design vectors in a model. If dummy coding is used, the Pearson correlation coefficient becomes a point-biserial correlation whereas in the case of effect coding, the groups coded by 1 and -1 are included in the computation and the 0-groups excluded.

However, at this point it should be clear that it is admissible to compute correlations between design vectors and frequency vectors since these correlations reflect the corresponding contrasts. In addition, the logarithm of the frequency vector can be correlated as well; the amount of correlation changes but since the logarithm is monotonic increasing, the ordering of the correlation coefficients remains the same. And this property is important for dominance analysis, as will become clear in the following elaborations. For further reading it is referred to Mair (2006).

Back to dominance analysis: Since this approach has to be modified with respect to log-linear models, $\log(\mathbf{m})$ is used. First, Budescu (1993) states that a predictor $\mathbf{x}_i$ "weakly dominates" $\mathbf{x}_j$ (i.e. $\mathbf{x}_i$ is as least as important as $\mathbf{x}_j$) if the multiple correlations

$$r^2_{\log(\mathbf{m}).\mathbf{x}_i \mathbf{x}_h} \geq r^2_{\log(\mathbf{m}).\mathbf{x}_j \mathbf{x}_h} ,$$

where $\mathbf{x}_h$ stands for any subset of the remaining $p - 2$ variables (p design vectors in total).

An analogous form is

$$\left(r^2_{\log(\mathbf{m}).\mathbf{x}_i\mathbf{x}_h} - r^2_{\log(\mathbf{m}).\mathbf{x}_h}\right) \geq \left(r^2_{\log(\mathbf{m}).\mathbf{x}_j\mathbf{x}_h} - r^2_{\log(\mathbf{m}).\mathbf{x}_h}\right),$$

where the variable's usefulness in the context of dominance is regarded. The shorthand notation of dominance is

$$\mathbf{x}_i \mathscr{D} \, \mathbf{x}_j \, .$$

Therefore, $\mathbf{x}_i$ is said to weakly dominate $\mathbf{x}_j$, if in all subset models that do not include either of the two predictors, $\mathbf{x}_i$ is at least as useful as $\mathbf{x}_j$ (i.e., $\mathbf{x}_i$ contributes to the overall fit of the model as much as $\mathbf{x}_j$ does). The possible relationships for two predictors $\mathbf{x}_i$ and $\mathbf{x}_j$ are

- $\mathbf{x}_i \mathscr{D} \, \mathbf{x}_j$
- $\mathbf{x}_j \mathscr{D} \, \mathbf{x}_i$
- $\mathbf{x}_i \mathscr{D} \, \mathbf{x}_j$ and $\mathbf{x}_j \mathscr{D} \, \mathbf{x}_i$
- $\mathbf{x}_i \mathscr{D} \, \mathbf{x}_j$ or $\mathbf{x}_j \mathscr{D} \, \mathbf{x}_i$.

The third case above defines equal importance and the fourth case indicates situations where the relative importance cannot be determined.

At this point we know about the dominance of the predictors. But the initial question was to determine a value for relative contribution of single effects independent from the other effects in the model. From the explanations above Budescu (1993) proposes, based on a work by Kruskal (1987), to compute the average usefulness (importance) $C^{(k)}_{\mathbf{x}_i}$ of $\mathbf{x}_i$ across all $\binom{p-1}{k}$ models consisting of $k + 1$ variables. The corresponding predictor subset is denoted by $\mathbf{x}_h$ and as a consequence, the importance measure is defined as

$$C^{(k)}_{\mathbf{x}_i} = \frac{\sum \left(r^2_{\log(\mathbf{m}).\mathbf{x}_i\mathbf{x}_h} - r^2_{\log(\mathbf{m}).\mathbf{x}_h}\right)}{\binom{p-1}{k}} \, .$$

Finally, the mean importance of $\mathbf{x}_i$ is

$$C_{\mathbf{x}_i} = \frac{1}{p} \sum_{k=0}^{p-1} C^{(k)}_{\mathbf{x}_i} \, .$$

An important point of the results given in Budescu (1993) and Kruskal (1987) is that by summing up over all predictors, an alternative way to compute a Pseudo-R^2 in log-linear models is accomplished. It is given by

$$R^2_{dom} = \sum_{i=1}^{p} C_{\mathbf{x}_i} \, .$$

When determining the relative importance, the R^2 acts as reference value in the sense that it is the total amount of explanation in the frequency vector achieved through the effects in the model. By dividing $C_{\mathbf{x}_i}/R^2$ for all $i = 1, \ldots, p$ the relative mean importance and contribution of each effect to model fit can be determined and it is denoted by $\psi_{rel}(\mathbf{x}_i)$.

To conclude, with the presented approach and in combination with the content based interpretability of the effects in the design matrix the single components (design vectors) of a standard as well as nonstandard log-linear model are completely analyzed.

2.3   Comparing Non-Nested Models

Log-linear models which are nested within each other are comparable through a LR-statistic. In the non-nested case, model comparison is usually performed on a descriptive level by comparing some information criteria like AIC or BIC. There are several procedures that allow to test two non-nested models against each other (Linhart, 1988; Vuong, 1989; Greene, 2003). For the current log-linear model analysis framework, Linhart's procedure is used. The reason for applying this method which is briefly discussed in this section, is that it was already successfully used for the comparison of hierarchical log-linear models (Linhart and Zucchini, 1986). However, there are no simulation studies or theoretical comments concerning the performance of the mentioned tests with respect to log-linear modeling that the author is aware of.

In Linhart's test the operating distribution function is given by $F(x)$ and the approximating distribution function by $G_\theta(x)$; $\theta \in \Theta \subset \mathbb{R}^p$, $x \in \mathbb{R}^k$. The basis for the test statistic is the Kullback-Leibler (KL) discrepancy

$$\Delta(\theta) = -\mathrm{E}_F\left(\log g_\theta(x)\right) ,$$

where g_θ is the corresponding density. The distribution functions of the two models to be compared are denoted by $G^{(1)}_{\theta^{(1)}}$ and $G^{(2)}_{\theta^{(2)}}$. Under certain regularity conditions the following hypothesis about the expected discrepancies can be stated:

$$\mathrm{E}\left(\Delta^{(1)}(\hat{\theta}^{(1)}_n)\right) \leq \mathrm{E}\left(\Delta^{(2)}(\hat{\theta}^{(2)}_n)\right) .$$

In other words this hypothesis states that model 1 fits significantly better than model 2 (the smaller the KL discrepancy the better the model fits). In order to achieve a test statistic the joint asymptotic distribution of $\sqrt{n}(\Delta^{(i)}_n(\hat{\theta}^{(i)}_n) - \Delta^{(i)}(\theta^{(i)}_0))$, for $i \in \{1,2\}$, has to be determined. By using the central limit theorem it follows that

$$\begin{pmatrix} \sqrt{n}\left(\Delta^{(1)}_n(\theta^{(1)}_0) - \Delta^{(1)}(\theta^{(1)}_0)\right) \\ \sqrt{n}\left(\Delta^{(2)}_n(\theta^{(2)}_0) - \Delta^{(2)}(\theta^{(2)}_0)\right) \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, \Lambda) .$$

Note that due to asymptotic reasons $\hat{\theta}^{(i)}_n$ is substituted by $\theta^{(i)}_0$ (see also Jennrich, 1969). The elements λ_{ij} (with $i, j \in \{1,2\}$) of the variance-covariance matrix Λ are

$$\lambda_{ij} = \mathrm{E}_F\left(\log g^{(i)}_{\theta^{(i)}_0}(x) \log g^{(j)}_{\theta^{(j)}_0}(x)\right) - \mathrm{E}_F\left(\log g^{(i)}_{\theta^{(i)}_0}\right) \mathrm{E}_F\left(\log g^{(j)}_{\theta^{(j)}_0}\right) .$$

It follows that

$$\sqrt{n}\frac{(\Delta^{(1)}_n(\hat{\theta}^{(1)}_n) - \Delta^{(2)}_n(\hat{\theta}^{(2)}_n)) + (\Delta^{(1)}(\theta^{(1)}_0) - \Delta^{(2)}(\theta^{(2)}_0))}{\sqrt{\lambda_{11} + \lambda_{22} - 2\lambda_{12}}} \xrightarrow{\mathcal{D}} N(0, 1) .$$

Linhart and Zucchini (1986) show that for the expected discrepancies the following approximation holds asymptotically:

$$\Delta^{(i)}(\theta^{(i)}_0) \approx \mathrm{E}_F\left(\Delta^{(i)}_n(\hat{\theta}^{(i)}_n)\right) + \frac{p^{(i)}}{2n} .$$

Table 2: Data example.

	Male		Female	
	Small Networks	Large Networks	Small Networks	Large Networks
Married	48	87	5	14
Nonmarried	78	45	130	109

Under the hypothesis that the two expected discrepancies are equal it follows that the components in the numerator of the equation above are Akaike's criteria for both models, i.e. $AIC^{(1)}$ and $AIC^{(2)}$. Finally, Linhart (1988) proposes the test statistic

$$z_{AIC} = \sqrt{n} \frac{AIC^{(1)} - AIC^{(2)}}{\sqrt{\lambda_{11} + \lambda_{22} - 2\lambda_{12}}} \, ,$$

which is asymptotically $N(0,1)$-distributed. This asymptotic distribution remains unchanged if λ_{ij} is replaced by $\hat{\lambda}_{ij}$, i.e. $\hat{\theta}_n^{(i)}$ instead of $\theta_0^{(i)}$ in the equation for λ_{ij}. For contingency tables usually $AIC = G^2 + 2p$. If the given test is applied then the numerator in z_{AIC} changes to

$$AIC^{(1)} - AIC^{(2)} = \frac{(G^2(M^{(1)}) + 2p^{(1)}) - (G^2(M^{(2)}) + 2p^{(2)})}{2n} \, ,$$

where $p^{(1)}$ and $p^{(2)}$ are the number of free parameters in $M^{(1)}$ and $M^{(2)}$, respectively (see Linhart, 1988, p. 160). An application example of this test statistic is given in the next section and for further elaborations it is referred to Shimodaira (1997).

3   Algorithmical Implementation and Data Example

At this point some work flow explanations with respect to a programming implementation in R are given. The corresponding flow chart of the whole routine is given in Figure 1. The input of this routine are the frequency vector and the design matrix.

Based on this matrix, a model is build up by starting from the base model $M^{(0)}$ and by adding gradually design vectors until a certain model is achieved. After determining the fit and the relative importance of the effects in the model, another model can be specified and the procedure starts from the beginning. If these two models are nested, they are compared with a LR-statistic whereas in the non-nested case z_{AIC} is computed. At the end, both models are analyzed and tested against each other.

The following example is taken from von Eye and Niedermeier (1999) and the corresponding data are given in Table 2. The frequencies correspond to a cross-classification of Marital Status (M; 1 = married, 2 = not married), Gender (G; 1 = male, 2 = female), and Size of Social Network (S; 1 = small, 2 = large). In total there are $N = 516$ subjects. This simple example shows strikingly the usefulness of nonstandard log-linear models in terms of imposing special hypotheses and improving the model fit. The corresponding
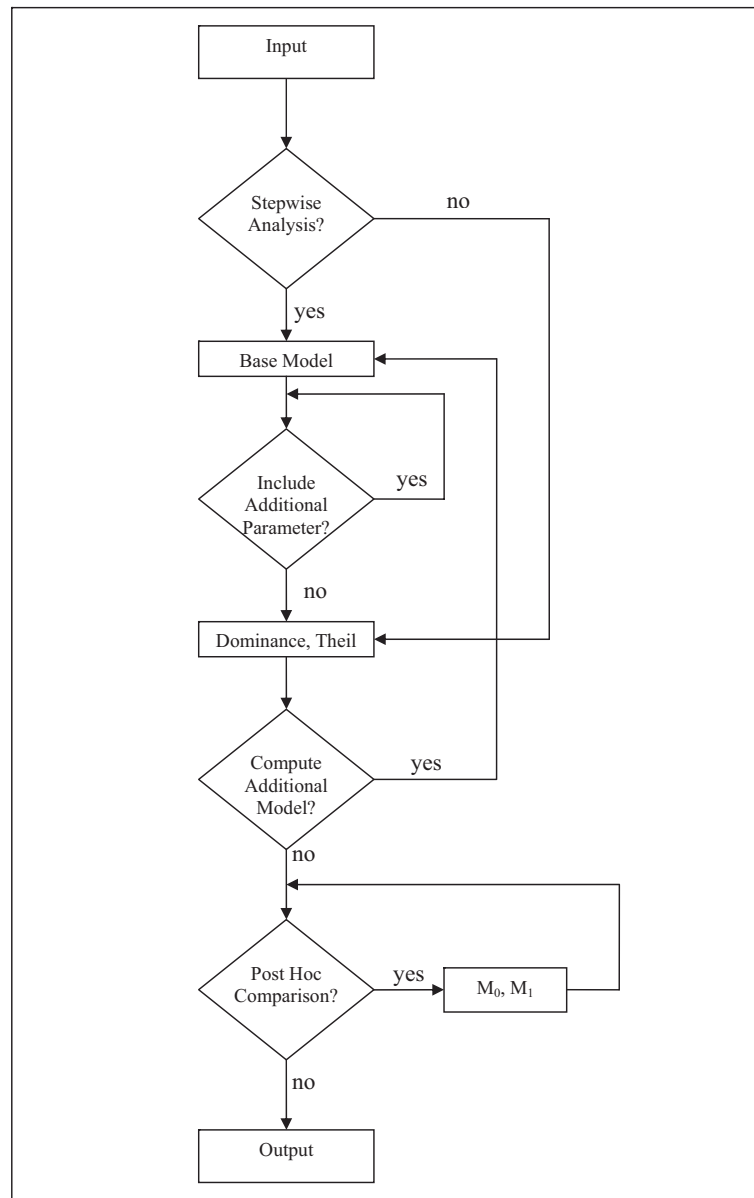
Figure 1: Flowchart of the Framework

GLM representation of the log-linear model is

$$
\log \begin{pmatrix} m_{111} \\ m_{112} \\ m_{121} \\ m_{122} \\ m_{211} \\ m_{212} \\ m_{221} \\ m_{222} \end{pmatrix} = \log \begin{pmatrix} 48 \\ 87 \\ 5 \\ 14 \\ 78 \\ 45 \\ 130 \\ 109 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 0 \\ 1 & -1 & 1 & -1 & -1 & 0 \\ 1 & -1 & -1 & 1 & 1 & 0 \\ 1 & -1 & -1 & -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}.
$$

Table 3: Results of stepwise analysis.

Model	df	Deviance	R^2_{MF}	AIC
$\log(\mathbf{m}) = \beta_0 \mathbf{1}$	7	249.21	—	296.28
$\log(\mathbf{m}) = \beta_0 + \beta_4 \mathbf{x}_4$	6	141.07	0.43	190.14
$\log(\mathbf{m}) = \beta_0 + \beta_4 \mathbf{x}_4 + \beta_1 \mathbf{x}_1$	5	54.79	0.78	105.86
$\log(\mathbf{m}) = \beta_0 + \beta_4 \mathbf{x}_4 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2$	4	26.68	0.89	79.75
$\log(\mathbf{m}) = \beta_0 + \beta_4 \mathbf{x}_4 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_5 \mathbf{x}_5$	3	11.47	0.95	66.53
$\log(\mathbf{m}) = \beta_0 + \beta_4 \mathbf{x}_4 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_5 \mathbf{x}_5 + \beta_3 \mathbf{x}_3$	2	3.38	0.98	60.45

This model is nonstandard since the last effect is a special coding vector and, thus, the design matrix deviates from an ordinary effect or dummy coding. The last vector is the translation of the assumption that among married people there is a difference in the size of social networks and this statement can be regarded as a special hypothesis. Moreover, the first vector is the intercept, the next three are the main effects and the third is the interaction $[M : G]$.

By using the presented stepwise approach to build up the model and by starting with the intercept only model $\log(\mathbf{m}) = \beta_0 \mathbf{1}$, the effects are carried in the order as given in Table 3.

To determine the goodness-of-fit, the deviance, McFadden's R^2_{MF}, and AIC are computed. The final model fits the data very well with $R^2_{MF} = 0.98$ and the parameters are entered in the following order: $\mathbf{x}_4$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_5$, and $\mathbf{x}_3$. This issue is confirmed by the dominance analysis, where the dominance ordering is given by

$$\mathbf{x}_4 \,\mathscr{D}\, \mathbf{x}_1 \,\mathscr{D}\, \mathbf{x}_2 \,\mathscr{D}\, \mathbf{x}_5 \,\mathscr{D}\, \mathbf{x}_3 \,.$$

These results suggest that the most important vector with respect to model fit is the interaction term $[M : G]$, followed by the main effects for M and G. Finally, the special coding vector dominates the main effect S. Basically, this approach can be considered as an ordinary analysis of deviance for log-linear models where the gathering of the parameters do not follow any hierarchical order.

From a content based view the question is whether the hypothesis stated through the coding vector is substantial (model $M^{(1)}$) with respect to, e.g., a main effects model $M^{(0)}$. The corresponding LR-statistic is

$$G^2(M^{(0)} \mid M^{(1)}) = G^2(M^{(0)}) - G^2(M^{(1)}) = 162.86 - 139.64 = 24.04\,.$$

The $df = 4 - 3 = 1$ and, thus, a significant p-value results which suggests that the improvement in the model due to the special effect is substantial.

Furthermore, using the dominance analysis the mean importances of the single vectors in the model context are computed. The corresponding relative mean importances of the effects are

$$\psi_{rel}(\mathbf{x}_1) = 0.377\,, \quad \psi_{rel}(\mathbf{x}_2) = 0.104\,, \quad \psi_{rel}(\mathbf{x}_3) = 0.013\,,$$
$$\psi_{rel}(\mathbf{x}_4) = 0.430\,, \quad \psi_{rel}(\mathbf{x}_5) = 0.077\,.$$

Therefore, a $R^2_{dom} = 0.99$ results. When comparing R^2_{dom} to $R^2_{MF} = 0.98$ as computed before and it is obvious that both R^2-values are almost equal.

Next, by using the same frequency table, a second model is fitted in order to apply Linhart's test procedure. This model and the former one are non-nested. The GLM equation is of the form

$$
\log
\begin{pmatrix}
m_{111} \\
m_{112} \\
m_{121} \\
m_{122} \\
m_{211} \\
m_{212} \\
m_{221} \\
m_{222}
\end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & -1 & 1 & -1 & -1 \\
1 & 1 & -1 & 1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 \\
1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix} .
$$

Instead of the special contrast vector, this model contains the remaining two-way interactions $[M : S]$ and $[G : S]$ and is denoted as M^2; the former model is named as $M^{(1)}$. The question of interest is whether $M^{(2)}$ fits significantly better than the more parsimonious $M^{(1)}$. Thus, the corresponding hypothesis can be stated as

$$
\begin{aligned}
H_0 &: \ AIC^{(2)} \geq AIC^{(1)} \\
H_1 &: \ AIC^{(2)} < AIC^{(1)}
\end{aligned}
$$

H_1 corresponds to the research question above. Using the equations given in Section 2.3, it follows that $\lambda_{11} = 0.290$, $\lambda_{12} = 0.285$, and $\lambda_{22} = 0.287$; $AIC^{(1)} = 60.45$, and $AIC^{(2)} = 59.08$. Therefore,

$$
z_{AIC} = \sqrt{516} \frac{(60.45 - 59.08)/1032}{\sqrt{0.290 + 0.287 - 2 \times 0.285}} = 0.36 \,,
$$

which corresponds to a p-value of 0.36 in the $N(0, 1)$-distribution. Thus, the higher parameterized model $M^{(2)}$ does not provide a significantly better fit than the more parsimonious model $M^{(2)}$ with less parameters. This is an important issue in nonstandard log-linear modeling: Such models can have less parameters but fit satisfactorily.

4 Discussion

Nonstandard log-linear models are a flexible tool to analyze categorical data. The estimation of these models is straightforward (ML), whereas the interpretation of the parameters can be rather ambiguous. In this paper, an analysis framework is presented that focuses on the model effects. Initially, design vectors are included stepwise. These components of the design matrix are interpretable formally. From a numerical point of view there was a lack in parameter interpretability with respect to the contribution to model fit. The solution elaborated here adopts the concept of dominance analysis to log-linear models. The key question that can be answered at the end is whether a certain design vector $\mathbf{x}_i$ is more or less (or equally) important than $\mathbf{x}_j$ in the context of the effects included in the selected model. The higher the importance, the higher the contribution of an effect to model fit. In addition, by summing up over the single effect importances, a method results to compute

an R^2 in log-linear models. How the corresponding R^2_{dom} performs in comparison to other Pseudo-R^2 in categorical data analysis, has not been studied yet.

Finally, to complete the framework, Linhart's test is used to compare two non-nested models. This test statistic is based on the comparison of the AIC's of the corresponding models. As mentioned, several procedures exist to test models on a non-nested level. To determine which test statistic is the most suitable, further studies are needed. However, the current procedure provides an analysis tool for standard and nonstandard log-linear models with respect to the mentioned topics. A corresponding source code in R is available upon request.

Acknowledgements

References

Alba, R. D. (1988). Interpreting the parameters in log-linear models. In J. S. Long (Ed.), *Common Problems/Proper Solutions* (p. 258-287). Newbury Park, CA: Sage.

Azen, R., and Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*, 129-148.

Benedetti, J. K., and Brown, M. P. (1978). Strategies for the selection of log-linear models. *Biometrics*, *34*, 680-686.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Bock, R. D. (1975). *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill.

Breen, R. (1984). Fitting nonhierarchical and association log-linear models using GLIM. *Sociological Methods and Research*, *13*, 77-107.

Brown, M. P. (1976). Screening effects in multidimensional contingency tables. *Applied Statistics*, *25*, 37-46.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542-551.

Edwards, D., and Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika*, *72*, 339-351.

Elliott, G. C. (1988). Interpreting higher order interactions in log-linear analysis. *Psychological Bulletin*, *103*, 121-130.

Evers, M., and Namboodiri, N. K. (1979). On the design matrix strategy in the analysis of categorical data. *Sociological Methodology*, *10*, 86-111.

Eye, A. von, and Niedermeier, K. E. (1999). *Statistical Analysis of Longitudinal Categorical Data in the Social and Behavioral Sciences*. Mahwah, NJ: Erlbaum.

Eye, A. von, and Spiel, C. (1996). Standard and nonstandard log-linear symmetry model for measuring change in categorical variables. *The American Statistician*, *50*, 300-305.

Fienberg, S. (1979). A note on fitting and interpreting parameters in models for categorical data. *Sociological Methodology*, *10*, 112-118.

Gilbert, G. N. (1981). *Modeling Society: An Introduction to Log-Linear Models for Social Researchers*. Boston, MA: George Allen and Unwin.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, *59*, 319-352.

Goodman, L. A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, *13*, 33-61.

Greene, W. H. (2003). *Econometric Analysis* (5 ed.). London: Prentice Hall.

Holt, D. (1979). Log-linear models for contingency table analysis: On the interpretation of parameters. *Sociological Methods and Research*, *7*, 330-336.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, *40*, 633-643.

Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician*, *41*, 6-10.

Linhart, H. (1988). A test whether two AIC's differ significantly. *South African Statistical Journal*, *22*, 153-161.

Linhart, H., and Zucchini, W. (1986). *Model Selection*. New York: Wiley.

Magidson, J., Swan, J. H., and Berk, R. A. (1981). Estimating nonhierarchial and nested log-linear models. *Sociological Methods and Research*, *10*, 3-49.

Mair, P. (2006). *Interpreting Standard and Nonstandard Log-Linear Models*. New York, Münster: Waxmann.

Mair, P., and Eye, A. von. (in preparation). Application scenarios for nonstandard log-linear models.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). London, New York: Chapman and Hall.

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, *135*, 370-384.

Rindskopf, D. (1990). Nonstandard log-linear models. *Psychological Bulletin*, *108*, 150-162.

Rindskopf, D. (1999). Some hazards of using nonstandard log-linear models, and how to avoid them. *Psychological Methods*, *4*, 339-347.

Rudas, T. (1988). *Odds Ratios in the Analysis of Contingency Tables*. Thousand Oaks, CA: Sage.

Shimodaira, H. (1997). Assessing the error probability of the model selection test. *Annals of the Institute of Statistical Mathematics*, *49*, 395-410.

Shtatland, E. S., Kleinman, K., and Cain, E. M. (2002). One more time about r^2 measures of fit in logistic regression. In *Proceedings of Northeast SAS User Groups (NESUG 15)*. Buffalo, NY.

Vermunt, J. K. (1997). *LEM: A General Program for the Analysis of Categorical Data*. Tilburg University.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*, 307-333.

Wickens, T. P. (1989). *Multiway Contingency Tables Analysis for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Wilson, T. P. (1979). On not interpreting coefficients: Comment on Holt. *Sociological Methods and Research*, 8, 233-240.

Author's address:

Patrick Mair
Department of Statistics and Mathematics
Vienna University of Economics and Business Administration
Augasse 2-6
A-1090 Vienna

Tel. ++43 1 31336 4844
Fax ++43 1 31336 774
E-mail: `patrick.mair@wu-wien.ac.at`