A NOTE ON APPLYING THE BCH METHOD UNDER LINEAR EQUALITY AND INEQUALITY CONSTRAINTS

L. Boeschoten

tilburg university & statistics netherlands

M.A. Croon

tilburg university

D.L. Oberski

utrecht university

May 10, 2017

Correspondence should be sent to

E-Mail: l.boeschoten@tilburguniversity.edu
Phone: +31(0)615167469
Address: PO box 90153, 5000 LE Tilburg

A NOTE ON APPLYING THE BCH METHOD UNDER LINEAR EQUALITY AND INEQUALITY CONSTRAINTS

Abstract

Researchers often wish to relate estimated scores on latent variables to exogenous covariates not previously used in analyses. The BCH method has been shown to correct for asymptotic bias in estimates due to these scores' uncertainty and has been shown to be relatively robust. When applying the BCH approach however, two problems arise. First, negative cell proportions can be obtained. Second, the approach cannot deal with situations where marginals need to be fixed to specific values, such as edit restrictions. The BCH approach can handle these problems when placed in a framework of quadratic loss functions and linear equality and inequality constraints. This research note gives the explicit form for equality constraints and demonstrates how solutions for inequality constraints may be obtained using numerical methods.

Key words: Classification, Latent class analysis, Three-step procedure, BCH method

Introduction

Researchers in many different disciplines apply latent structure models in which observed variables are treated as indicators of an underlying latent variable that can not be measured directly. An often used strategy in this context consists of three steps (Vermunt, 2010). First, the parameters of the measurement model are estimated, describing the relationship between the latent variable and its indicators. Second, each respondent is assigned a latent score based on his/her scores on the indicators. Finally, the relationships between the latent scores and scores on exogenous variables are assessed.

Croon (2002) showed that for general latent structure models such a strategy leads to inconsistent estimates of the parameters of the joint distribution of the latent and the exogenous variables. Bolck, Croon & Hagenaars (2004) discussed this problem in the context of latent class analysis where observed variables are categorical. They also derived a correction procedure that produces consistent estimates, known as the BCH correction method. Subsequent simulation studies by Vermunt (2010), Bakk, Tekle & Vermunt (2013), Bakk & Vermunt (2016) and Nylund-Gibson & Masyn (2016) have demonstrated that this procedure produces unbiased parameter estimates and correct inference in a large range of simulations. When applying the BCH correction method, two problems can arise. First, negative cell proportion estimates can be obtained (Asparouhov & Muthén, 2015). Second, the approach cannot deal with situations where marginals need to be constrained. An example is edit restrictions in official statistics, which leads to certain marginals being fixed to zero (De Waal, Pannekoek & Scholtus, 2012).

In this research note we extend the BCH method to solve these two problems. We allow for linear equality and inequality constraints by noting the correction method minimizes a quadratic loss function and give a closed form solution for linear equality restrictions. Next, we demonstrate how solutions for inequality constraints may be obtained using numerical methods. We first discuss the three-step approach to the latent class model and the BCH correction method. We then show how to impose linear restriction and how to extend this to including non-negativity constraints. In the appendix, we give R code to apply the procedure.

The three-step approach to the latent class model and the BCH correction method

Let us denote a set of observed exogenous variables $\mathbf{Q}$ and an unobserved latent variable $\mathbf{X}$. In the context of latent class models, all variables involved are assumed to be categorical. Let $\mathbf{Q} = (\mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_J)$ be the Cartesian product of J different discrete random variables $\mathbf{Q}_j$. If the variable $\mathbf{Q}_j$ is defined for n_j categories, the distribution of $\mathbf{Q}$ can be specified as a multinomial distribution with $n = \prod_{j=1}^{J} n_j$ categories.

In the basic latent class model considered by Bolck, Croon & Hagenaars (2004), a single categorical latent variable $\mathbf{X}$ with m categories is introduced. The variable $\mathbf{X}$ itself is not directly observed but only indirectly via a set of indicator variables $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_K)$. Let the joint distribution of the categorical variables $\mathbf{Q}$, $\mathbf{X}$ and $\mathbf{Y}$ be denoted by

$$\mathrm{p}(\mathbf{Q} = \mathbf{q}, \mathbf{X} = x, \mathbf{Y} = \mathbf{y}) = \mathrm{p}(\mathbf{q}, x, \mathbf{y}).$$

Then a possible factorization is

$$\mathrm{p}(\mathbf{q}, x, \mathbf{y}) = \mathrm{p}(\mathbf{q})\mathrm{p}(x|\mathbf{q})\mathrm{p}(\mathbf{y}|x, \mathbf{q}).$$

Since in the basic latent class model $\mathbf{Q}$ is assumed to have no direct effect on $\mathbf{Y}$, the latter result simplifies to

$$\mathrm{p}(\mathbf{q}, x, \mathbf{y}) = \mathrm{p}(\mathbf{q})\mathrm{p}(x|\mathbf{q})\mathrm{p}(\mathbf{y}|x).$$

The three-step approach to the estimation of the parameters of the latent class model starts with the estimation of the parameters of the measurement model represented by the conditional probability distribution $\mathrm{p}(\mathbf{y}|x)$. Once this estimation procedure is completed, individual research units may be assigned to one of the latent classes solely on the basis of their observed scores on $\mathbf{Y}$. This defines the second step of the estimation procedure and results in an assignment of each individual to a latent class. If the random variable $\mathbf{W}$ represents the latent classes individuals are assigned to, and assignment is done using a modal rule where each individual is assigned to the class for which its posterior membership probability is the largest, this can be expressed as

$$\mathrm{p}(w|\mathbf{y}) = \begin{cases} 1 & \text{if } \mathrm{p}(x_1|\mathbf{y}) > \mathrm{p}(x_2|\mathbf{y}) \ \forall\ x_1 \neq x_2, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

Different assignment rules than the modal rule will yield a different form for Equation 1. All subsequent results also apply to other assignment rules, such as proportional or random assignment (Bakk, 2015).

Since $\mathbf{Y}$ and $\mathbf{Q}$ are conditionally independent given $\mathbf{X}$, so are $\mathbf{W}$ and $\mathbf{Q}$ and the conditional distributions are related by

$$\mathrm{p}(w|\mathbf{q}) = \sum_{x=1}^{\mathbf{X}} \mathrm{p}(w|x)\mathrm{p}(x|\mathbf{q}).$$

In terms of the joint distribution this becomes

$$\mathrm{p}(\mathbf{q}, w) = \sum_{x=1}^{\mathbf{X}} \mathrm{p}(\mathbf{q}, x)\mathrm{p}(w|x).$$

The latter result can be recast as a matrix equation

$$\mathbf{E} = \mathbf{AD},$$

with the elements of the three matrices defined as $e_{qw} = \mathrm{p}(q, w)$, $a_{qx} = \mathrm{p}(q, x)$ and $d_{xw} = \mathrm{p}(w|x)$. After completing the first and the second estimation step, the elements of the matrices $\mathbf{E}$ and $\mathbf{D}$ are known. The joint distribution of $\mathbf{Q}$ and the latent variable $\mathbf{X}$ is then given by

$$\mathbf{A} = \mathbf{ED}^{-1}.$$

Here it is assumed that matrix $\mathbf{D}$ is not singular so that its inverse exists. See Bolck, Croon & Hagenaars (2004, pp.13–14) for a discussion on when this assumption may be violated. A consistent estimate of $\mathbf{A}$ is $\hat{\mathbf{E}}\hat{\mathbf{D}}^{-1}$.

The previously obtained algebraic solution for matrix $\mathbf{A}$ can also be derived via a rather trivial minimization of a least squares function. Let $\mathbf{E}$ and $\mathbf{D}$ be matrices with known elements. Matrix $\mathbf{E}$ is of order $n \times m$ and $\mathbf{D}$ is an invertible matrix of order $m \times m$. Let $\mathbf{A}$ be an $n \times m$ matrix of unknown elements and consider the following least squares function:

$$\varphi = \frac{1}{2}\mathrm{tr}(\mathbf{AD} - \mathbf{E})'(\mathbf{AD} - \mathbf{E}).$$

Minimizing φ with respect to the unknown matrix $\mathbf{A}$ yields $\mathbf{A} = \mathbf{ED}^{-1}$, for which φ attains the truly minimal value of zero. Note that the factor $1/2$ is introduced to obtain simpler expressions for the first derivatives. Its introduction does not change the solution of the minimization problem.

The correction procedure under linear equality constraints

In some applications, simple linear restrictions may be imposed on the elements of matrix $\mathbf{A}$. For instance, some of the probabilities in the joint distribution of $\mathbf{Q}$ and $\mathbf{X}$ may be set equal to zero, for example for combinations of $\mathbf{Q}$ and $\mathbf{X}$ that cannot occur in practice. After imposing such zero constraints, all the non-zero cell probabilities should still add to one. The quadratic loss function φ can be minimized under equality constraints on the unknown elements of matrix $\mathbf{A}$ by applying the method of Lagrangian multipliers.

We first rewrite the quadratic loss function φ in the following way using vectorization operations on matrices, see Schott (1997, pp. 261–266). For the vector of residuals $\mathbf{r}$ we obtain

$$\begin{aligned}\mathbf{r} &= \mathrm{vec}(\mathbf{AD} - \mathbf{E}) \\ &= \mathrm{vec}(\mathbf{I}_{n\times n}\mathbf{AD}) - \mathrm{vec}(\mathbf{E}),\end{aligned}$$

where $\mathbf{I}_{n\times n}$ is an $n \times n$ identity matrix. Applying Theorem 7.15 from Schott (1997, p.263) yields

$$\mathbf{r} = (\mathbf{D}' \otimes \mathbf{I}_{n\times n}) \cdot \mathrm{vec}(\mathbf{A}) - \mathrm{vec}(\mathbf{E}),$$

in which $\otimes$ is the Kronecker product of two matrices (Graham, 1982). Defining $\mathbf{P} = \mathbf{D}' \otimes \mathbf{I}_{n\times n}$, $\mathbf{a} = \mathrm{vec}(\mathbf{A})$ and $\mathbf{e} = \mathrm{vec}(\mathbf{E})$, we are able to write

$$\mathbf{r} = \mathbf{Pa} - \mathbf{e},$$

so that the least squares function becomes

$$\begin{aligned}\varphi &= \frac{1}{2}\mathbf{r}'\mathbf{r} \\ &= \frac{1}{2}(\mathbf{a}'\mathbf{P}'\mathbf{Pa} - 2\mathbf{e}'\mathbf{Pa} + \mathbf{e}'\mathbf{e}).\end{aligned}$$

The completely unconstrained solution to the minimization problem is given by

$$\mathbf{a}_0 = (\mathbf{P}'\mathbf{P})^{-1} \cdot \mathbf{P}'\mathbf{e}.$$

Now suppose that the S linear equality constraints can be represented by a matrix equation

$$\mathbf{Ha} = \mathbf{c}.$$

The matrix $\mathbf{H}$ is of order $S \times N$, N being the number of cells in matrices $\mathbf{A}$ and $\mathbf{E}$. We may assume that $\mathbf{H}$ is of rank S; otherwise, the linear equality constraints would not be linearly independent. In order to minimize the least square function φ under a set of S linear constraints on the elements of $\mathbf{A}$, the Lagrangian is defined as

$$\mathbf{L} = \varphi - \lambda'(\mathbf{Ha} - \mathbf{c}). \tag{2}$$

Setting the first derivatives of $\mathbf{L}$ with respect to $\mathbf{a}$ equal to the zero vector, and solving for $\mathbf{a}$ yields:

$$\mathbf{a} = (\mathbf{P}'\mathbf{P})^{-1}(\mathbf{P}'\mathbf{e} + \mathbf{H}'\lambda),$$

which can be rewritten as:

$$\mathbf{a} = \mathbf{a}_0 + (\mathbf{P}'\mathbf{P})^{-1}\mathbf{H}'\lambda.$$

Solving for the unknown Lagrangian multipliers by taking the derivative of the Lagrangian (equation 2), and setting it to zero, or equivalently by imposing linear constraints $\mathbf{Ha} - \mathbf{c} = 0$ yields:

$$\lambda = [\mathbf{H}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{H}']^{-1}(\mathbf{c} - \mathbf{Ha}_0).$$

So that the final solution for $\mathbf{a}$ is:

$$\mathbf{a} = \mathbf{a}_0 + (\mathbf{P}'\mathbf{P})^{-1}\mathbf{H}'[\mathbf{H}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{H}']^{-1}(\mathbf{c} - \mathbf{Ha}_0).$$

Note that the vector $\mathbf{c} - \mathbf{Ha}_0$ represents the deviations of the unconstrained solution from the linear equality constraints. Again a consistent estimate of $\mathbf{a}$ can be obtained by replacing $\mathbf{P}$ and $\mathbf{a}_0$ with their sample estimates.

The correction procedure under linear equality and inequality constraints

A second issue with the BCH procedure is that in finite samples the consistent estimate $\hat{\mathbf{A}}$ hat may contain negative values. This issue is similar to the occurrence of Heywood cases in factor analysis (Heywood, 1931). Such negative values in the probability table estimate $\hat{\mathbf{A}}$ may prevent subsequent analyses. We suggest to prevent such inadmissible solutions by imposing inequality constraints. The resulting minimization problem is a quadratic program that can be solved by an iterative method.

Such a numerical iterative method for an equality and inequality constrained minimization of a quadratic function has been described by Goldfarb & Idnani (1983). Their numerical algorithm solves the quadratic programming problem of the form

$$\min(\frac{1}{2}\mathbf{b}'\mathbf{D}_{\text{mat}}\mathbf{b} - \mathbf{d}'_{\text{vec}}\mathbf{b}),$$

subject to the constraints

$$\mathbf{H}'\mathbf{b} \geq \mathbf{b}_0,$$

with respect to the n unknown parameters in vector $\mathbf{b}$. The matrix $\mathbf{D}_{\text{mat}}$ is a given $n \times n$ symmetric positive definite matrix whereas $\mathbf{d}_{\text{vec}}$ is a given $n \times 1$ vector.

In order to apply the Goldfarb-Idnani optimization procedure in the present context, the following definitions have to be implemented. First, to include non-negativity constraints, we make use of Theorem 7.6 from Schott (1997, p.254) to obtain

$$\begin{aligned}\mathbf{D}_{\text{mat}} &= \mathbf{P}'\mathbf{P} \\ &= (\mathbf{D}\mathbf{D}') \otimes \mathbf{I}_{n \times n}.\end{aligned}$$

and

$$\mathbf{d}_{\text{vec}} = \mathbf{e}'\mathbf{P}$$

Since it is assumed that matrix $\mathbf{D}$ is of full rank, the matrix $\mathbf{P}'\mathbf{P}$ is positive-definite. This ensures that the quadratic loss function φ is strictly convex. Moreover, the type of equality and inequality constraints considered here (the sum of the elements in matrix $\mathbf{A}$ is equal to 1, where all elements ≥ 0 and some are fixed to 0), define a convex region in the parameter space.

In order to represent the constraints on the cell probabilities we now define matrix $\mathbf{H}$ in such a way that the first row of $\mathbf{H}$ has all its elements equal to 1. This row represents a constraint on the sum of all cell probabilities. We represent this row vector as matrix $\mathbf{H}_0$. Let $J = \{1, 2, 3, ..., N\}$ be an index set corresponding to the column numbers of matrix $\mathbf{H}$. This index set can be partitioned in two non-overlapping subsets J_1 and J_2:

- Subset J_1 contains the indices of the elements of vector $\mathbf{a}$ which are set exactly equal to zero: for those indices j we require $\mathbf{a}_j = 0$;
- Subset J_2 contains the indices of the elements of vector $\mathbf{a}$ which are required to be non-negative: for those indices j we require $\mathbf{a}_j \geq 0$.

Now let I_n be an $N \times N$ identity matrix and permute the rows of this matrix so that the upper part contains the rows corresponding with the index numbers in J_1, and the lower part of the permuted identity matrix contains the rows corresponding with the index numbers in J_2. Referring to the two parts of the permuted identity matrix as $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively, the matrix $\mathbf{H}$ is obtained by

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix},$$

where $\mathbf{H}$ is used to obtain the final solution for $\mathbf{a}$. Note that in cases where we are not interested in applying equality constraints, but we are interested in applying the inequality constraints we simply omit $\mathbf{H_1}$. Vector $\mathbf{b_0}$ is of length $N + 1$, with its first element equal to 1 and all the remaining elements equal to 0.

With this procedure, we are able to find a solution for $\mathbf{A}$ (the joint distribution of latent variable $\mathbf{X}$ and exogenous covariates $\mathbf{Q}$) where the sum of the elements is equal to 1, where no negative elements are created, and where impossible combinations of scores can be set to have a probability of zero. Having defined $\mathbf{b}$, $\mathbf{D}_{\text{mat}}$ and $\mathbf{H}$, the solution can be obtained using standard software for quadratic programming, such as the R package `quadprog` (Turlach & Weingessel, 2013).

Conclusion

We have modified the BCH procedure to include linear equality and inequality constraints solving the problem of negative solutions and allowing for restrictions on arbitrary cell margins. With these adjustments, analysts interested in relating covariates to assignments on latent class variables will now be able to, for example, impose edit restrictions, further analyse solutions that were previously inadmissible and analyse datasets involving more complex marginal restrictions. The appendix gives R code demonstrating the application of our method to an example with linear inequality constraints.

Appendix

The iterative method for an equality and inequality constrained minimization of a quadratic function described by Goldfarb & Idnani (1983) has been implemented in the R package `quadprog` available in the repository CRAN (Turlach & Weingessel, 2013).

The minimization procedure is implemented in the function `solve.QP` which is called as

```
solve.QP(Dmat,dvec,Amat,bvec,meq).
```

Its arguments are:

- `Dmat`: the matrix $\mathbf{D}$ appearing in the quadratic function: $(\mathbf{D}\mathbf{D}') \otimes \mathbf{I}_{n\times n}$;
- `dvec`: the vector $\mathbf{d}$ appearing in the quadratic function: $\mathbf{e}'\mathbf{P}$;
- `Amat`: The transpose of $\mathbf{H}$ ($\mathbf{H}'$) defining the linear constraints on the parameters $\mathbf{b}$;
- `bvec`: A vector of length $N+1$, with its first elements equal to 1 and the reminaing N elements all equal to 0, these are the constants $\mathbf{b}_0$ in the constraints.
- `meq`: 1+ the number of elements in J_1

Suppose we have a variable with five categories, imputed by using a latent class model, and an exogenous variable with six categories and we can specify the $\mathbf{E}$-matrix as:

$$\mathbf{E} = \begin{pmatrix} 0.1447 & 0.3511 & 0.1245 & 0.1369 & 0.2428 \\ 0.1689 & 0.3205 & 0.1534 & 0.1615 & 0.1957 \\ 0.1864 & 0.2836 & 0.1832 & 0.1846 & 0.1622 \\ 0.2972 & 0.2445 & 0.2133 & 0.2057 & 0.0394 \\ 0.3020 & 0.2063 & 0.2430 & 0.2243 & 0.0244 \\ 0.3015 & 0.1711 & 0.2721 & 0.2404 & 0.0149 \end{pmatrix}$$

and the $\mathbf{D}$-matrix as:

$$\mathbf{D} = \begin{pmatrix} 0.9038576 & 0.02625164 & 0.02983604 & 0.03463155 & 0.005423212 \\ 0.0938927 & 0.894957 & 0 & 0.0111503 & 0 \\ 0.1235493 & 0 & 0.8221187 & 0.0006061352 & 0.05372586 \\ 0.4043909 & 0.04771681 & 0.004035212 & 0.521612 & 0.02224529 \\ 0.06167063 & 0 & 0.3387299 & 0.01368246 & 0.585917 \end{pmatrix}$$

The unconstrained solution for the **A**-matrix can be obtained by:

```
A <- E%*%solve(D)
```

```
> A
              [,1]      [,2]        [,3]      [,4]         [,5]
[1,] -0.014252890 0.3796841 -0.01684810 0.2446348  0.406782024
[2,] -0.009650054 0.3426973  0.05450798 0.2945295  0.317915255
[3,] -0.010903852 0.2990010  0.11722723 0.3414568  0.253218792
[4,]  0.097073516 0.2500092  0.24182405 0.3815053  0.029687906
[5,]  0.085326726 0.2056358  0.29115309 0.4196604 -0.001776127
[6,]  0.069457951 0.1649944  0.33578910 0.4529588 -0.023200287
```

It is immediately clear that this solution contains negative values and is therefore inadmissible. If we use our proposed solution, we start by specifying which combinations of scores between **X** and **Q** are in practice not possible. If, for example, the combinations $X = 2; Q = 6$ and $X = 3; Q = 6$ are not possible, we need to specify that in (vector) **a**, cells 12 and 18 are going to be fixed to 0:

```
iequal <- c(12,18)
```

Next, we are making use of the following function:

```
qpsolve <- function(e,d,iequal){
  nr    <- nrow(e)
  nc    <- ncol(e)
  ncel  <- nr*nc
  evec  <- as.vector(e)
```

```
  id    <- diag(nr)
  p     <- kronecker(t(d),id)
  dmat  <- kronecker(d %*% t(d),id)
  dvec  <- as.vector(evec %*% p)
  im    <- diag(ncel)
  i1    <- iequal
  i2    <- setdiff(1:ncel,i1)
  index <- c(i1,i2)
  im2   <- im[index,]
  at    <- rbind(rep(1,ncel),im2)
  amat  <- t(at)
  bvec  <- c(1,rep(0,ncel))
  meq   <- 1 + length(iequal)
  res   <- solve.QP(dmat,dvec,amat,bvec,meq)
  return(res)
}
```

We then use the function by defining the **E**-matrix, the **D**-matrix and the inequality constraints:

```
res <- qpsolve(E,D,iequal)
```

We see here that the unconstrained solution is equal to the unconstrained solution we previously obtained:

```
> matrix(res$unconstrained.solution, ncol=5)
             [,1]      [,2]        [,3]      [,4]         [,5]
[1,] -0.014252890 0.3796841 -0.01684810 0.2446348  0.406782024
[2,] -0.009650054 0.3426973  0.05450798 0.2945295  0.317915255
[3,] -0.010903852 0.2990010  0.11722723 0.3414568  0.253218792
[4,]  0.097073516 0.2500092  0.24182405 0.3815053  0.029687906
[5,]  0.085326726 0.2056358  0.29115309 0.4196604 -0.001776127
[6,]  0.069457951 0.1649944  0.33578910 0.4529588 -0.023200287
```

The constrained solution looks as follows:

```
> round(matrix(res$solution, ncol=5),9)
          [,1]       [,2]       [,3]       [,4]       [,5]
[1,] 0.00000000 0.17368232 0.00000000 0.00000000 0.01212123
[2,] 0.00000000 0.14310137 0.00000000 0.00000000 0.00000000
[3,] 0.00000000 0.10467234 0.00000000 0.00000000 0.00000000
[4,] 0.10599014 0.05879943 0.01421970 0.01279514 0.00000000
[5,] 0.09500016 0.01453725 0.04917607 0.04872426 0.00000000
[6,] 0.09329787 0.00000000 0.00000000 0.07388273 0.00000000
```

We see here that no negative values are produced, that cells $X = 2; Q = 6$ and $X = 3; Q = 6$ are exactly zero and we see that

```
> sum(matrix(res$solution, ncol=5))
[1] 1
```

It is interesting to note that if a researcher is not dealing with the problem of specific cells that need to be fixed to 0, but is interested in applying the inequality constraints, this argument can simply be left empty:

```
iequal <- c()
```

References

Asparouhov, T. & Muthén, B. (2015). Auxiliary variables in mixture modelling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes 21*. Version was retrieved April 26th, 2017 from https://www.statmodel.com/examples/webnotes/webnote21.pdf

Bakk, Z. (2015). *Contributions to bias adjusted stepwise latent class modeling* (Doctoral thesis, Tilburg University, Tilburg, The Netherlands). Retrieved from https://pure.uvt.nl/portal/files/8521154/Bakk_Contributions_16_10_2015.pdf

Bakk, Z., Tekle, F.B., & Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, *43*, 272–311. http://dx.doi.org/10.1177/0081175012470644

Bakk, Z., & Vermunt, J.K. (2016). Robustness of stepwise latent class modelling with continuous distal outcomes. *Structural Equation Modelling: A Multidisciplinary Journal*, *23*, 20–31. http://dx.doi.org/10.1080/10705511.2014.955104

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*, 3–27. http://dx.doi.org/10.1093/pan/mph001

Croon, M. (2002). Using predicted latent scores in general latent structure models. In I. Marcoulides G.A. & Moustaki (Eds.), *Latent variable and latent structures models*(pp. 195–224). Mahwah NJ: Lawrence Erlbaum.

De Waal, T., Pannekoek, J., & Scholtus, S. (2012). The editing of statistical data: methods and techniques for the efficient detection and correction of errors and missing values. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*, 204–210. http://dx.doi.org/10.1002/wics.1194

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming*, *27*, 1–33. http://dx.doi.org//10.1007/BF02591962

Graham, A. (1982). *Kronecker products and matrix calculus: With applications.* New York, NY: John Wiley & Sons Inc.

Heywood, H.B. (1931) On finite sequences of real numbers. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *134*, 486–501. http://dx.doi.org//10.1098/rspa.1931.0209

Nylund-Gibson, K & Masyn, K.E. (2016) Covariates and Mixture Modeling: Results of a Simulation Study Exploring the Impact of Misspecified Effects on Class Enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 782–797. http://dx.doi.org/10.1080/10705511.2016.1221313

Schott, J. R. (1997). *Matrix analysis for statistics.* New York, NY: Wiley & Sons Inc.

Turlach, B. A., & Weingessel, A. (2013). Quadprog: Functions to solve quadratic programming problems. *R package version 1.5–5.* Version was retrieved May 1^{st}, 2017 from https://cran.r-project.org/web/packages/quadprog/quadprog.pdf

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political analysis*, *18*, 450–469. dx.doi.org//10.1093/pan/mpq025