

Stepwise Latent Class Models for Explaining Group-Level Outcomes Using Discrete
Individual-Level Predictors

Margot Bennink, Marcel A. Croon, and Jeroen K. Vermunt

Tilburg University

Abstract

Explaining group-level outcomes from individual-level predictors requires aggregating the observed scores on these predictors to the group level and accounting for the measurement error in the aggregated scores to prevent biased estimates. However, it is not clear how to perform the aggregation and the correction for measurement error when the individual-level predictors are discrete variables. It is shown how to overcome these problems by a stepwise latent class analysis. In the first step, a latent class model is estimated in which the scores on a discrete individual-level predictor are used as indicators for a group-level latent class variable. In the second step, this latent class model is used to aggregate the individual-level predictor to the group-level by assigning the groups to the latent classes. In the final step, a group-level analysis is performed in which the aggregated measures are related to the remaining group-level variables while correcting for the measurement error in the class assignments. The proposed stepwise model is compared to existing methods in a simulation study and extended to a situation with multiple group-level latent variables. In the end, the approach is applied to an empirical data example in which team productivity (group level) is explained by job control (individual level), job satisfaction (individual level), and enriched job design (group level).

Keywords: micro-macro analysis, latent class analysis, multilevel analysis, discrete variables, stepwise modeling

Stepwise Latent Class Models for Explaining Group-Level Outcomes Using Discrete Individual-Level Predictors

Though typically multilevel models attempt to explain an individual-level dependent variable by means of individual- and group-level predictors (Goldstein, 2011; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012), the prediction of group-level outcomes is equally important. Snijders and Bosker (2012) refer to the latter type of multilevel analysis as micro-macro analysis since a micro (or individual)-level predictor is assumed to affect a macro (or group)-level outcome. Such a micro-macro analysis is relevant, for example, when a health psychologist is interested in whether students' attitudes (micro-level predictor) affect teacher's stress (macro-level outcome), or when a developmental psychologist wishes to investigate whether parenting practices (macro-level predictor) affect their children's study habits (micro-level mediator), which in turn affects children's achievement (micro-level mediator), which subsequently affects parental stress (macro-level outcome) (Bovaird & Shaw, 2012). These micro-macro relationships cannot be addressed within the mainstream multilevel framework (Preacher, Zyphur, & Zhang, 2010).

Traditionally, data from micro-macro designs are analyzed by aggregation; that is, the group means of the individual-level variables are assigned to the groups, and subsequently a group-level analysis is performed using the aggregated individual scores and group-level variables. Note that this is in fact a stepwise procedure since the aggregation (measurement model) is separated from the group-level analysis (structural model). An example from group-performance research is provided by van Veldhoven (2005), who studied the relationships between perceived human resource practices, work climate, and job stress on the one hand, and prospective and retrospective financial performance on the other hand. Because the financial performance indicators are only available at the business level, individual survey scores were aggregated to mean scores to perform a single-level analysis at the business level.

Although very intuitive, the above aggregation approach has various serious drawbacks. One is that it implicitly assumes that the group members provide perfect

information about their group, while in practice it is more realistic to expect that the group means contain measurement error. Croon and van Veldhoven (2007) showed that ignoring this measurement error causes a bias in the estimates of the parameters from the structural model. Another problem of the aggregation approach is that it is not clear how to aggregate categorical predictors to the group level. For example, for nominal variables with more than two categories, the group mean has no substantive meaning. It might be more appropriate to use the group modes instead, but even then the problem of ignoring measurement error remains.

The current article presents how it is possible to keep working in the same stepwise matter, so separating the aggregation from the group-level analysis, but correct the group-level estimates for the bias that is caused by the measurement error in the aggregated scores. To overcome the measurement error issue, a latent variable model for two-level data is used, in which the individual-level responses serve as indicators for a latent variable at the group-level. Then this latent group-level variable is used as a predictor of the group-level outcome variable (Croon & van Veldhoven, 2007). It is also possible to perform a micro-macro multilevel analysis with categorical variables in this way by using a latent class model with a categorical latent variable at the group-level.

A short description of the stepwise procedure is that, in the first step, a latent class model is estimated in which the scores on the discrete individual-level predictor are used as indicators for a group-level latent class variable (measurement model). In the second step, this latent class model is used to aggregate the individual-level predictor to the group-level by assigning the groups to the latent classes. In the final step, a group-level analysis is performed in which the aggregated measures are related to the remaining group-level variables (structural model), while adjusting for the measurement errors in the class assignments. The latter adjustments are based on earlier work by Bolck, Croon, and Hagenars (2004); Vermunt (2010); Bakk, Tekle, and Vermunt (2013), . This stepwise latent class approach to micro-macro analysis with discrete data makes it possible to make adjustments in the structural model without changing the measurement model on which the aggregation is based. So when a group-level variable is

added or removed from the group-level analysis, the assigned or aggregated group-level scores remain unchanged. This would not be the case when the measurement and structural model are estimated simultaneously in a one-step procedure.

The organization of the article is as follows. First, the latent class model for discrete micro-macro analysis is introduced using a model that contains a single individual-level variable. Second, it is shown how this model can be applied in a stepwise manner. Third, a simulation study is presented to evaluate the performance of the proposed stepwise procedure. Fourth, the stepwise procedure is applied to a more complex model with two individual-level variables and, last, applied to a real data example in which enriched job design (macro-level predictor) affects team productivity (macro-level outcome) directly, and indirectly through job control (micro-level mediator) and job satisfaction (micro-level mediator).

Multilevel Latent Class Model for Micro-Macro Analysis

To illustrate the multilevel latent class model for micro-macro analysis, let us consider a simple model that contains a single group-level outcome Y_j , a single group-level predictor X_j , and a single individual-level predictor Z_{ij} . Subscript j is used to denote a particular group and subscript i to denote the individuals within a group. The group-level predictor is expected to affect the group-level outcome directly and indirectly via the individual-level predictor. Any theory in which a group-level intervention is not only expected to influence a group-level (performance) measure directly, but also indirectly through a characteristic of the group members, can be tested with this model. These kinds of models are sometimes referred to as 2-1-2 models since the effect of the level-2 independent variable on the level-2 dependent variable is mediated by a level-1 variable. Also, the term "bathtub model" is in use here because of the shape of the conceptual model that is shown in Figure 1. In this conceptual model, the latent variable is presented in a circle and the manifest variables in rectangles.

Figure 1 about here

The model of interest is a latent class model for two-level data in which the scores

of the individual-level units i within group j on the micro-level predictor Z_{ij} are treated as indicators of a discrete latent class variable defined at the group-level, ζ_j . Thus, the number of indicators of the latent variable equals the number of individuals within a group. This part of the model is referred to as the measurement part. All group members are treated as equivalent sources of information about the group-level variable; therefore, no one is considered as providing more accurate judgments in this respect than his co-members. This implies that the relationship between the individual-level variable and the group-level latent variable can be assumed to be the same for all individuals within a group. According to the local independence assumption commonly made in latent class analysis, the individual responses of group members are independent given the score of their group on the latent variable. In the structural part of the model, the group-level latent classes are related to the group-level independent variable X_j and the group-level dependent variable Y_j .

For the general case where all variables in the latent class model are discrete, the model can be formally described with three multi-category logit models (Agresti, 2013). Let there be P , Q , R , and S nominal response categories for, respectively, Z_{ij} , ζ_j , X_j , and Y_j and a particular category is denoted by p , q , r , and s . Then there are $P - 1$, $Q - 1$, and $S - 1$ different logit equations defined for, respectively, Z_{ij} , ζ_j , and Y_j in which each category is compared to an arbitrary baseline category. When the first categories are used as baselines, the multinomial logit equations are:

$$\log \left(\frac{P(Z_{ij} = p | \zeta_j = q)}{P(Z_{ij} = 1 | \zeta_j = q)} \right) = \beta_p^Z + \beta_{p q}^{Z \zeta}, \quad (1)$$

$$\log \left(\frac{P(\zeta_j = q | X_j = r)}{P(\zeta_j = 1 | X_j = r)} \right) = \beta_q^\zeta + \beta_{q r}^{\zeta X}, \quad (2)$$

and

$$\log \left(\frac{P(Y_j = s | \zeta_j = q, X_j = r)}{P(Y_j = 1 | \zeta_j = q, X_j = r)} \right) = \beta_s^Y + \beta_{s q}^{Y \zeta} + \beta_{s r}^{Y X}. \quad (3)$$

Each equation contains an intercept term (β_p^Z , β_q^ζ , and β_s^Y), and an effect for each of the predictor variables ($\beta_{p q}^{Z \zeta}$, $\beta_{q r}^{\zeta X}$, $\beta_{s q}^{Y \zeta}$, $\beta_{s r}^{Y X}$).

Equation 1 describes the measurement part of the model in which the scores of the individual-level units on the micro-level predictor Z_{ij} are treated as exchangeable indicators of a discrete latent class variable defined at the group-level, ζ_j . The structural part of the model is defined in Equations 2 and 3 in which the group-level latent variable is related to the other group-level variables: ζ_j is regressed on X_j , and the outcome Y_j is regressed on ζ_j and X_j .

The parameters of this latent class model can simultaneously be estimated by full information maximum likelihood estimation and, therefore, this method is further referred to as the one-step approach. This approach mainly has two drawbacks. First, it is not intuitive to simultaneously aggregate the individual-level variable to the group-level and relate the aggregated scores to the remaining variables since researchers are used to work in a stepwise manner when a manifest mean or mode is used in the group-level analysis. Second, the definition of the latent group-level variable is not only determined by the micro-level indicators, but also by the remaining variables in the structural model. Thus, when the structural part of the model is adapted, say a level-2 covariate or outcome is added or removed, the full model has to be re-estimated and the measurement model may thus change. Especially, the fact that the meaning and possibly also the number of the latent classes is affected by the outcome variable is very undesirable since the latent classes were theoretically intended to predict this outcome. These problems associated with the simultaneous estimation of the model are circumvented with a stepwise approach.

Stepwise Estimation

A stepwise estimation procedure of the micro-macro latent class model consists of the following three steps:

- First step: Estimate the measurement model (i.e., relate the micro-level predictor to the latent group-level variable).
- Second step: Aggregate the micro-level predictor to the group-level by assigning groups to latent classes.

- Third step: Estimate the structural part of the model while correcting for the classification errors that were made in the second step.

So, as graphically illustrated in Figure 2(a), a measurement model is defined for ζ_j and the corresponding latent class model is estimated in the first step of the analysis:

$$P(\mathbf{Z}_j) = \sum_{q=1}^Q P(\zeta_j = q) \prod_{i=1}^{I_j} P(Z_{ij} | \zeta_j = q) . \quad (4)$$

Here, the vector \mathbf{Z}_j contains the I_j responses Z_{ij} of the members of group j . The model parameters are the class proportions $P(\zeta_j = q)$ and the conditional response probabilities $P(Z_{ij} | \zeta_j = q)$, which, as shown in Equation 1, are typically parametrized using a logit equation. As in the one-step model, the responses of the individual group members on Z_{ij} are assumed to be exchangeable and independent given the latent classes, but the meaning of the latent classes is now only determined by the individual-level scores on the micro-level predictor and no longer by the scores on the group-level variables. The number of latent classes of ζ_j is also determined during this step.

Figure 2 about here

In the second step, the groups are assigned to the Q latent classes based on their scores \mathbf{Z}_j . We denoted the assigned class for group j by W_j . The assignment process in which the new variable W_j is constructed is graphically illustrated in Figure 2(b). As in a standard latent class analysis, the class assignments are obtained using the posterior class membership probabilities $P(\zeta_j = q | \mathbf{Z}_j)$ from the first step. Several types of deterministic and probabilistic assignment rules have been proposed, but in the current article, we focus on modal and proportional assignment. With modal assignment, each group is assigned to the latent class for which the posterior probability is largest. With proportional assignment, a group is assigned to each of the Q classes with a probability equal to the posterior membership probability for the class concerned.

Unless the micro-level predictor is a perfect indicator for the group-level latent class variable, classification errors will be made during the assignment. In order to account for these classification errors (in step three), we use the $Q \times Q$ classification

table with entries $P(W_j = t|\zeta_j = q)$. Note that $P(W_j = t|\zeta_j = q)$ is the conditional probability that a group belonging to class q is assigned to class t of W_j ($t = 1, \dots, Q$). The off-diagonals of this table represent classification error probabilities. In Appendix A, we show how these probabilities can be obtained from the latent class model parameters.

In the third step, the structural model is estimated; that is, the assigned scores W_j are related to the other group-level variables, in this case X_j and Y_j . As shown by Bolck et al. (2004) biases are caused by the classification errors introduced in the second step; that is, by the fact that we have W_j instead of ζ_j . However, they also indicated that it is possible to adjust for the classification errors, which prevent biases. Key for their adjustment method is the following relationship between $P(Y_j, W_j|X_j)$ and $P(Y_j, \zeta_j|X_j)$:

$$\begin{aligned} P(Y_j, W_j|X_j) &= \sum_{q=1}^Q P(Y_j, \zeta_j = q|X_j)P(W_j|\zeta_j = q) \\ &= \sum_{q=1}^Q P(\zeta_j = q|X_j)P(Y_j|X_j, \zeta_j = q)P(W_j|\zeta_j = q). \end{aligned} \quad (5)$$

This equation shows that a model for $P(Y_j, \zeta_j|X_j)$ is obtained by estimating a model for $P(Y_j, W_j|X_j)$ and correcting for the probabilities $P(W_j|\zeta_j = q)$. Note that the $P(W_j|\zeta_j = q)$ were computed in step two. The resulting model is shown graphically in Figure 2(c).

The model defined in Equation 5 can be estimated by maximum likelihood (ML). This involves performing a latent class analysis in which W_j is used as a single indicator, and in which the probabilities $P(\zeta_j = q|X_j)$ and $P(Y_j|X_j, \zeta_j = q)$ and the corresponding logit coefficients (see Equation 2 and 3) are freely estimated, and in which $P(W_j|\zeta_j = q)$ is treated as known and thus fixed. We refer to this approach as the ML 3-step method (Bakk et al., 2013; Vermunt, 2010). An alternative proposed by Bolck et al. (2004) – and that we therefore call the BCH 3-step approach – involves reformulating the problem into a weighted analysis (see also Vermunt (2010)). More specifically, by weighting the data points by the inverse of the classification probabilities $P(W_j|\zeta_j)$, we adjust for the fact that the W_j contain classification errors. The reweighted data can be used as observed data to estimate the structural parameters of interest.

From previous research on three-step latent class analysis, it is known that the bias adjusted stepwise approaches work very well in situations encountered in practice. However, this approach may fail when a small sample size is combined with a very large proportion of classification errors, where the latter can also be quantified as class separation (Bakk et al., 2013; Vermunt, 2010). In such situations, the maximum likelihood estimates of the step-1 latent class model will tend to yield classes being more different than they truly are (Galindo-Garre & Vermunt, 2006). Consequently, the amount of classification errors is underestimated and the structural parameters are not adjusted to a sufficient degree.

The proportion of classification errors and the class separation is mainly a function of the number of indicator variables (in micro-macro analysis, this equals the number of individuals within each group), the number of classes, and the response probabilities for the most likely response. Since all group members are assumed to be exchangeable, the response probabilities for the most likely response are the same for individuals within the same group, which makes the measurement model of a micro-macro model more parsimonious than a regular latent class analysis. The following simulation study, we investigate under which conditions class separation is large enough to perform an unbiased stepwise analysis for the current model.

Simulation Study

In this section, the performance of the stepwise approaches is first evaluated and compared to manifest aggregation with a group mode, one-step latent aggregation, and stepwise latent aggregation without correcting for measurement error. Second, the lower boundary of the separation between classes is explored at which still unbiased results are obtained with the stepwise approaches. All analyses were carried out with Latent GOLD 5.0 (Vermunt & Magidson, 2013).

Data are generated according to the model shown in Figure 1 with all dichotomous variables. An average situation is created by fixing the between-effects from the structural part of the model to .4 on a logistic scale using effect coding

($\beta^{YX} = \beta^{\zeta X} = \beta^{Y\zeta} = .4$). The number of groups was fixed to 100 ($J = 100$) and the number of individuals within a group to 10 ($n_j = 10$) which are minimum sample sizes for this type of analysis (Bennink, Croon, & Vermunt, 2013). It is expected that larger samples provide slightly better results. The relationship between the scores on the micro-level predictor and the latent variable is varied from weak to strong, again using effect coding ($\beta^{Z\zeta} = .2, .4, .6, \text{ or } .8$). This corresponds to class separations, measured with the entropy based R-square, of $R_{entr}^2 = .24, .64, .88, \text{ and } .97$. For each of the four conditions, 500 datasets were generated and analyzed with manifest mode aggregation, the latent variable one-step approach, and the latent variable three-step approaches. The three-step procedures were applied with both modal and proportional assignment.¹

Table 1 about here

The average estimated structural parameters in each condition are presented in Table 1 and should be compared with their true value of .4. As expected, the one-step and the corrected three-step procedures provide unbiased results, regardless whether the modal or proportional assignment rule was used. Only when the quality of the indicators is extremely poor ($\beta^{Z\zeta} = .2$) do both methods perform less well, which, as shown below, is the situation in which classes are estimated as being more different than they truly are. Furthermore, as can be seen, the standard deviations of the estimates decrease when the quality of the indicators increases.

The other methods fail. When the uncorrected three-step method is used, the estimates of the group-level relationships in which ζ_j is involved are underestimated in line with Bolck et al. (2004), regardless whether the modal or proportional assignment rule was used. Because the indirect effect is underestimated, the direct effect is overestimated. The bias decreases when the strength of the relationship between Z_{ij} and ζ_j increases. When Z_{ij} is aggregated to the group-level using the manifest group

¹In the one-step and the three-step ML methods, we used weakly informative priors for the model probabilities to prevent boundary estimates for the logit parameters. Because this does not work in the three-step BCH method, 82 datasets with the modal assignment rule and 73 datasets with the proportional assignment rule did not converge and were excluded from the further analysis.

mode, the estimates of the between-parameters are biased and this bias decreases when the quality of the indicators improves. In line with previous research (Bennink et al., 2013), the parameter estimates are only unbiased when the strength of the relationship between the individual-level predictor and the corresponding group-level variable is very good ($\beta^{Z\zeta} = .8$). The standard deviations of the estimates obtained with mode aggregation and uncorrected three-step are stable over the true quality of the indicators.

Table 2 about here

As shown above, the corrected three-step approaches perform less well with poor indicators, the situation corresponding with an extremely low class separation of .24.² To illustrate why this occurs, the true and estimated proportion of classification errors are shown in Table 2 for each condition. As can be seen, the proportion of classification errors is underestimated in the condition in which $R_{entr}^2 = .24$ resulting in third-step parameters which are not sufficiently corrected. To explore what would be a sufficiently high class separation, also indicators with $\beta^{Z\zeta}$ values of .25, .30, and .35, corresponding to R_{entr}^2 values of .35, .45, and .55, are added to the table. As can be seen, with a class separation of .45, the estimated proportion of classification errors gets close to the actual proportion. This applies to both modal and proportional assignment. It can also be seen, that the variability (the standard deviation) of the estimated proportion of classification error decreases when the true quality of the indicators increases.

To conclude, the adjusted three-step approaches provide estimates that are as good as the estimates from a one-step analysis, as long as the parameters as the class separation is sufficient ($R_{entr}^2 = .45$). It does not matter whether a BCH or ML correction is used or whether modal or proportional assignment is used. When the relationship between the micro-level predictor and the group-level variable is very strong, all methods provide unbiased estimates, but this is not a realistic situation in

²As stated before, class separation is not only a function of the quality of the indicators but also of the number of indicators which in our application equals the number of individuals within a group. For example, an indicator with a $\beta^{Z\zeta}$ -value of .2 yields a $R_{entr}^2 = .42$ when $n_j = 20$ and $R_{entr}^2 = .55$ when $n_j = 30$.

practice. When the relationship between the micro-level predictor and the group-level variable is moderate, a latent variable should be used for the aggregation since the manifest mode aggregation provides biased estimates. When the relationship between the micro-level predictor and the group-level variable is extremely weak, all methods may provide biased estimates for the group-level sample size investigated in the current simulation study.

Multiple Macro-Level Latent Variables

The simple model discussed so far is extended to a model with two micro-level predictors, Z_{1ij} and Z_{2ij} , and thus two latent variables, ζ_{1j} and ζ_{2j} , as shown in Figure 3. The corresponding multinomial logit equations for Z_{1ij} and Z_{2ij} are the same as described in Equation 1, and the ones for ζ_{1j} and ζ_{2j} are the same as described in Equation 2.

Figure 3 about here

When X_j , ζ_{1j} , ζ_{2j} , and Y_j contain, respectively, R , W , Q , and S categories and a particular category is denoted by r , w , q , and s , the multinomial logit equation for Y_j is:

$$\log \left(\frac{P(Y_j = s | \zeta_{1j} = w, \zeta_{2j} = q, X_j = r)}{P(Y_j = 1 | \zeta_{1j} = w, \zeta_{2j} = q, X_j = r)} \right) = \beta_s^Y + \beta_{s w}^{Y \zeta_1} + \beta_{s q}^{Y \zeta_2} + \beta_{s r}^{Y X} . \quad (6)$$

The first categories are used as baseline categories, β_s^Y is the intercept of the response variable, and $\beta_{s w}^{Y \zeta_1}$, $\beta_{s q}^{Y \zeta_2}$, and $\beta_{s r}^{Y X}$ are the regression parameters of the predictor variables.

When the two individual-level predictors would be continuous variables, it would be common practice to include both their between- and within-group correlation in the model. In the case of discrete variables, this concept is translated by incorporating an association between Z_{1ij} and Z_{2ij} ($a_{Z_1 Z_2}$), and between ζ_{1j} and ζ_{2j} ($a_{\zeta_1 \zeta_2}$). Thus, Z_{1ij} and Z_{2ij} are not expected to be independent given the latent group-level variables. Instead, while keeping the latent group-level variables constant, there may still be some residual within-group association between the micro-level predictors. At the

between-level, it is also expected that there is some residual association between ζ_{1j} and ζ_{2j} , after controlling for X_j .

While estimating this model in a single step is straightforward, when estimating it stepwise, it has to be decided how to define the first step model(s). The first option would be to formulate a separate measurement model for ζ_{1j} and ζ_{2j} as described in Equation 4. By formulating two measurement models, the meaning of the latent classes is only influenced by the individual-level scores on the corresponding micro-level predictors. The number of latent classes for ζ_{1j} and ζ_{2j} can be determined without being influenced by the variables from the structural part of the model, but the eventual residual within-association among Z_{1ij} and Z_{2ij} is ignored.

An alternative would be to formulate a single simultaneous measurement model for the two latent variables which includes the residual within-association between Z_{1ij} and Z_{2ij} :

$$P(\mathbf{Z}_{1j}, \mathbf{Z}_{2j}) = \sum_{w=1}^W \sum_{q=1}^Q P(\zeta_{1j} = w, \zeta_{2j} = q) \prod_{i=1}^{I_j} P(Z_{1ij}, Z_{2ij} | \zeta_{1j} = w, \zeta_{2j} = q). \quad (7)$$

By doing so, the meaning of the latent classes is still not influenced by the group-level variables from the structural part of the model, but the number of latent classes for ζ_{1j} and ζ_{2j} should be determined simultaneously.

An analysis is carried out to explore whether the misspecification of the measurement model arising from ignoring the residual within-association among Z_{1ij} and Z_{2ij} affects the estimates of the between-level parameters. Since sampling fluctuation is not of primary interest here, one very large data set ($J = 10000, I_j = 100$) is generated in each of the investigated conditions. If it turns out that ignoring the within-association provides biased estimates in such very large samples, the estimates in smaller samples can be expected to be even worse because of sampling fluctuation.

The population models varied in the strength of the relationship between the latent variables and the corresponding micro-level predictors (indicators), and the strength of the within-association among the micro-level predictors (within-association):

- indicators ($\beta^{Z_1\zeta_1}$ and $\beta^{Z_2\zeta_2}$): 0.2, 0.4, or 0.6
- within-association ($a_{Z_1Z_2}$): 0, 0.2, 0.4, or 0.6

Similar to the previous simulation study, all variables, including the latent ones, are dichotomous and the between-effects from the structural part of the model are fixed to .4 on the logistic scale using effect coding

($\beta^{\zeta_1 X} = \beta^{\zeta_2 X} = a_{\zeta_1 \zeta_2} = \beta^{Y \zeta_1} = \beta^{Y \zeta_2} = \beta^{Y X} = .4$). A final remark is that because of the large number of individuals within a group, the R_{entr}^2 value is very high ($>.90$) in all conditions.

The generated datasets were analyzed with the one-step and the bias adjusted three-step approaches. In the one-step procedure, we used both the correct specification containing the residual within-association and the incorrect model excluding this association. In the first step of the BCH and ML bias adjusted stepwise procedures, we used either a single joint measurement model with the association between Z_{1ij} and Z_{2ij} or two separate measurement models which ignore this association. Both modal and proportional assignment rules were used to assign the groups to latent classes in the second step of the analysis. The manifest mode-aggregation and uncorrected three-step procedures were not used because the previous simulation study showed that these methods already fail in a simpler model.

Table 3 about here

Table 3 presents the results for the conditions in which $a_{Z_1 Z_2}$ is varied and $\beta^{Z_1 \zeta_1}$ and $\beta^{Z_2 \zeta_2}$ are fixed to .4. The reported results concern the between-level parameter which is most strongly affected by ignoring the within association; that is, the association between ζ_{j1} and ζ_{j2} ($a_{\zeta_1 \zeta_2}$), which has a true value of .4. As can be seen, when the within-association among Z_{1ij} and Z_{2ij} is correctly modeled, both the one-step and the bias adjusted three-step methods provide unbiased estimates. However, when this within-association is not taken into account, the between association estimate is biased with all estimation procedures. The larger the value of the ignored within-association among Z_{1ij} and Z_{2ij} , the more the between-association among ζ_{1j} and ζ_{2j} is overestimated. Note that the estimates obtained with the one-step and the various types of bias adjusted three-step estimates are all very similar. The estimates of the remaining between-effects are not as much biased as the between-association. When

there is bias in the remaining between-parameters it is a downwards bias that probably compensates the overestimation of the between-association among ζ_{1j} and ζ_{2j} .

Table 4 about here

Table 4 shows how the bias that is caused by ignoring the within-association among the micro-level predictors interacts with the quality of the micro-level scores as indicators for the group-level latent variables. With bad indicators and a strong ignored within-association ($\beta^{Z_1\zeta_1} = \beta^{Z_2\zeta_1} = 0.2$ and $a_{Z_1Z_2} = 0.6$), the estimates of the between-association among ζ_{1j} and ζ_{2j} are very biased, while with good indicators and a small ignored within-association ($\beta^{Z_1\zeta_1} = \beta^{Z_2\zeta_1} = 0.6$ and $a_{Z_1Z_2} = 0.2$) the estimates are still good.

All together, these results show that the bias adjusted three-step procedures can be used for this micro-macro model without introducing more bias compared to the one-step procedure, as long as the within-association among the micro-level predictors is modeled in the first step. The within-association can only be ignored when the micro-level scores are very good indicators of the group-level latent variables and the within-association is small.

Data example

The stepwise micro-macro model with two micro-level predictors is now applied to a real data example. Since the BCH and ML correction procedures and the modal and proportional assignment rules provided similar results in the simulation study, only one method is applied here, namely the ML three-step approach with modal assignment. The inspiration for the current data example comes from a paper by Croon, van Veldhoven, Peccei, and Wood (2014). These authors show the relevance of bathtub multilevel mediation models, such as the one discussed in this paper, for research on human resource management (more specifically on job design) and organizational performance by using an example from the Workplace Employment Relations Survey 2004 (WERS2004). This is a publicly available large-scale dataset from the United Kingdom with representative sampling at both the employee and the workplace level.

More information about the survey can be found at www.wers2004.info. Croon et al. (2014) investigated to what extent the relationship between the adoption of enriched job designs at the level of the workplace, on the one hand, and workplace labor productivity, on the other hand, was mediated, at the individual level, by employees' experienced sense of job control and job satisfaction. For the current application, all measures from Croon et al. (2014), *enriched job design*, *job control*, and *job satisfaction*, were categorized into variables with three categories of approximately equal size (low, medium and high), while *labor productivity* was transformed into a variable with two approximately equally sized categories (low and high). To keep the discussion simple, the current illustration ignores that the variables were originally measured with multiple items. The analyses were performed on 18,505 employees nested within 1,455 workplaces. More information about the Latent GOLD 5.0 (Vermunt & Magidson, 2013) syntax used in this analysis is given in Appendix B.

In the first step of the procedure, a measurement model is formulated in which the individual-level scores on *job control* (Z_{1ij}) and *job satisfaction* (Z_{2ij}) were used as exchangeable indicators for two latent variables at the group-level (ζ_{1j} and ζ_{2j}) with a residual within-association ($a_{Z_1Z_2}$) among the individual-level measures on *job control* and *job satisfaction* included as well. The optimal number of latent classes for the two latent variables was determined simultaneously by comparing fit indices of models with all possible combinations of one to five classes for each latent variable. Because the decision is about the number of classes at the group level, the fit indices that incorporate the sample size are based on the number of groups (Lukočienė, Varriale, & Vermunt, 2010). BIC, AIC3, CAIC, and SABIC were lowest for the model with three latent classes for both *job control* and *job satisfaction*. Contradictorily, AIC was lowest when both latent variables contained four latent classes. Since most fit indices point towards this direction and the individual-level observed variables had three response categories, the three-class solutions for both group-level latent variables were retained.

Table 5 about here

Table 5 displays the class sizes and the class-specific response probabilities for

both latent variables. These can be used to interpret the latent classes. Table 5(a) shows that the first latent variable has a class that contains 35% of the groups, and these workplaces contain mostly employees with a high probability of scoring low ($p = .70$) on *job control*. There is a second latent class that contains 56% of the groups, and in these workplaces, employees have the highest probability to score medium on *job control* ($p = .84$). The final class of workplaces contains 9% of groups that contain individuals that, like the ones from the second latent class, have the highest probability to score medium ($p = .66$), but they have a higher probability to score high ($p = .33$) on *job control*. From Table 5(b) it can be seen that the second latent variable has a class that contains 61% of the groups and these workplaces have the highest probability to score low on *job satisfaction* ($p = .88$). The second class contains 23% of the workplaces with employees having the highest probabilities to score medium ($p = .62$) on *job satisfaction*. The last class contains 16% of the workplaces in which employees have the highest probability to score high on *job satisfaction*. The groups from the third class also have quite a high probability to score low ($p = .34$), but this is probably caused by the fact that overall most groups score low on *job satisfaction* ($p = .69$). Based on these posterior probabilities, all workplaces are assigned to a particular latent class using a modal assignment rule. The class separation was .39 for the latent variable *job control* and .43 for *job satisfaction* which is relatively low for a three-step analysis.

Table 6 about here

In the third step, the assigned latent class variables were related to the group-level measures of *enriched job design* (X_j) and *labor productivity* (Y_j) in a latent class model in which the assigned class membership scores from the second step were used as single indicators with known measurement errors, namely the classification errors from the second step. The group-level parameters obtained with the ML bias adjusted three-step approach are presented in Table 6, from which the estimates of the parameters for the main effects of the response variables are omitted. For all effects, dummy coding was used with the first categories as reference categories. Only the significance of the global

effects are reported since the significance of the category-specific parameters depends on coding.

The overall effect of *enriched job design* on ζ_{1j} (job control) is significant ($\chi^2 = 23.71, df = 4, p < 0.001$) and the category-specific parameters are presented in Table 6(a). All category-specific parameters are positive; thus, the reference category scores lower than the other categories. $\beta_2^{\zeta_{1X}}$ and $\beta_3^{\zeta_{1X}}$ differ more from the reference category than $\beta_2^{\zeta_{2X}}$ and $\beta_3^{\zeta_{2X}}$. The overall effect of *enriched job design* on ζ_{2j} (job satisfaction) is also significant ($\chi^2 = 29.82, df = 4, p < 0.001$). The category-specific parameters from Table 6(b) show that the $\beta_2^{\zeta_{2X}}$ and $\beta_2^{\zeta_{3X}}$ are positive and $\beta_3^{\zeta_{2X}}$ and $\beta_3^{\zeta_{3X}}$ are negative. Hence, the first two categories score higher than the reference group and the latter two score lower than the reference group.

The overall association among ζ_{1j} and ζ_{2j} is significant ($\chi^2 = 13.92, df = 4, p = 0.008$), but the category-specific variables from Table 6(c) are not significant. It is, therefore, difficult to interpret the association.

The overall effect of ζ_{1j} (job control) on *labor productivity* is not significant ($\chi^2 = 1.37, df = 2, p = 0.51$), while the overall effects of ζ_{2j} (job satisfaction) and *enriched job design* are significant ($\chi^2 = 22.87, df = 2, p < 0.001$ and $\chi^2 = 4.97, df = 2, p = 0.08$), although the latter only at a significance level of 10% and not at 5%. The category-specific parameters are presented in Table 6(d). All categories except $\beta_2^{\zeta_{2Y}}$ score higher than the reference group.

To conclude, there is a significant direct effect of *enriched job design* on *labor productivity*. The two paths of the indirect effect of *enriched job design* (macro-level) on *labor productivity* (macro-level) through *job satisfaction* (micro-level) are significant, while only the first part of the indirect path through *job control* (micro-level) is significant and the second part is not significant.

Discussion

A stepwise multilevel latent class model was proposed to predict group-level outcomes by means of discrete individual- and group-level predictors. In the first step, a

latent class model was estimated in which the individual-level predictor was used as an indicator for a group-level latent class variable (measurement model). In the second step, the individual-level predictor was aggregated to the group level based on the latent class model from the first step. This had two important advantages. First, the measurement error in the aggregated scores is known. Second, it is an elegant way of aggregating a discrete variable since it is not very clear how to do this with a manifest mean or mode. Next, the aggregated scores are related to the remaining group-level variables while correcting for the known measurement error (structural model). It is shown that the bias adjusted stepwise procedures work without introducing bias since the results of the stepwise approaches were very similar to the parameters that are obtained when the measurement and structural model are simultaneously estimated in a one-step analysis. Since researchers are used to working in a stepwise manner when they aggregate with a manifest mode, they can continue to work in the way they are used to while accounting for measurement error in the aggregated scores and still get unbiased results.

Two issues are of importance. First, in case the model contains multiple macro-level latent variables, the within-association among the micro-level indicators needs to be included in the first step of the stepwise procedure, unless the within-association is small and the micro-level scores are very good indicators of the group-level latent variables. Conceptually, it would suit the philosophy of stepwise estimation better to formulate two separate measurement models in the first step, one for each latent variable, but the simulation study showed that ignoring this within-association provides biased estimates of the between-association among the latent variables. It is unrealistic to assume that there is no residual within-association among the predictors since that would imply that all associations among the micro-level predictors can be explained through the group-level latent variables. Second, class separation needs to be sufficiently high ($R_{entr}^2 = .45$) since results with poorly separated classes are only correct with large sample sizes. In practice, this is no problem, since it is of little use to aggregate a variable that will not or only weakly be

related to the group-level.

The stepwise ML procedure is applied to a real data example in which labor productivity (group-level outcome) is explained by enriched job design (group-level predictor), job control (individual-level predictor) and job satisfaction (individual-level predictor). All variables from the example were constructed from multiple items but were used as single variables in the model. For the continuous version of the current application, Croon et al. (2014) found that the factor analytic model was better equipped to detect bathtub-type linkages than a model using scale scores. A nice direction for further research would be to see whether this is also the case for discrete variables, thus look at micro-macro models in which especially the micro-level predictors, but also the group-level outcome and predictor, are latent variables measured with multiple items.

Last, in the current article, only attention is paid to the parameter estimates of the model and not to the standard errors of these parameters. Theory suggest that when fixed parameter estimates, obtained in the first step of the stepwise procedure, are plugged into the likelihood function, the effect of their sampling variability on the uncertainty about the estimates in the third step should be accounted for (Murphy & Topel, 1985). Fortunately, an easily accessible correction method for the standard errors is already made available by Bakk, Oberski, and Vermunt (accepted). In situations with a large sample size, like the current simulation study and data example, correction of the standard errors is not needed because the uncertainty about the estimates from the first step is very small (Bakk et al., accepted).

References

- Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Bakk, Z., Oberski, L. D., & Vermunt, J. K. (accepted). Relating latent class assignments to external variables: standard errors for correct inference. *Political analysis*.
- Bakk, Z., Tekle, F. T., & 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.
- Bennink, M., Croon, M. A., & Vermunt, J. K. (2013). Micro-macro multilevel analysis for discrete data: A latent variable approach and an application on personal network data. *Sociological Methods and Research*, *42*, 431-457.
- 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.
- Bovaird, J. A., & Shaw, L. H. (2012). Multilevel structural equation modeling. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (p. 501-518). New York, NY: The Guilford Press.
- Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model. *Psychological Methods*, *12*, 45-57.
- Croon, M. A., van Veldhoven, M. J. P. M., Peccei, R., & Wood, S. (2014). Researching individual well-being and performance in context: Multilevel mediational analysis for bathtub models. In M. J. P. M. van Veldhoven & R. Peccei (Eds.), *Well-being and performance at work: The role of context*. London, UK: Psychology Press.
- Galindo-Garre, F., & Vermunt, J. K. (2006). Avoiding boundary estimates in latent class analysis by bayesian posterior mode estimation. *Behaviormetrika*, *33*, 43-59.
- Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). West Sussex, UK: John Wiley & Sons, Inc.

- Hox, J. J. (2010). *Multilevel analysis: techniques and applications* (2nd ed.). New York, NY: Routledge.
- Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower and higher-level classes in multilevel latent class analysis. *Sociological Methodology, 40*, 247-283.
- Murphy, K. M., & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics, 3*, 88-97.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209-233.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage Publications.
- van Veldhoven, M. J. P. M. (2005). Financial performance and the long-term link with HR practices, work climate and job stress. *Human resource management journal, 15*, 30-53.
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*, 450-469.
- Vermunt, J. K., & Magidson, J. (2013). LG-syntax user's guide: Manual for latent GOLD 5.0 syntax module [Computer software manual]. Belmont, MA: Statistical Innovations Inc.

Table 1

Estimates Between-Effects Simple Micro-Macro Model

	Mode	1-step		3-step					
		none		BCH		ML			
		modal	prop	modal	prop	modal	prop		
Estimates of $\beta^{YX}(SD)^1$									
True $\beta^{Z\zeta} = .2$.49(.12)	.41(.19)	.49(.12)	.50(.11)	.45(.17)	.43(.18)	.45(.17)	.45(.13)	
True $\beta^{Z\zeta} = .4$.46(.12)	.41(.14)	.46(.12)	.47(.12)	.42(.14)	.41(.14)	.42(.14)	.42(.14)	
True $\beta^{Z\zeta} = .6$.42(.12)	.40(.13)	.42(.12)	.42(.12)	.41(.13)	.41(.13)	.41(.13)	.41(.13)	
True $\beta^{Z\zeta} = .8$.41(.13)	.41(.14)	.41(.13)	.41(.13)	.41(.14)	.41(.14)	.41(.14)	.41(.14)	
Estimates of $\beta^{\zeta X}(SD)^1$									
True $\beta^{Z\zeta} = .2$.18(.10)	.45(.25)	.19(.14)	.12(.08)	.36(.28)	.42(.26)	.37(.26)	.34(.18)	
True $\beta^{Z\zeta} = .4$.31(.11)	.42(.14)	.32(.11)	.28(.09)	.41(.15)	.41(.14)	.41(.16)	.41(.14)	
True $\beta^{Z\zeta} = .6$.40(.12)	.40(.12)	.37(.12)	.36(.11)	.40(.13)	.40(.12)	.40(.12)	.40(.12)	
True $\beta^{Z\zeta} = .8$.40(.11)	.41(.11)	.40(.11)	.40(.11)	.41(.11)	.41(.11)	.41(.11)	.41(.11)	
Estimates of $\beta^{\zeta Y}(SD)^1$									
True $\beta^{Z\zeta} = .2$.17(.12)	.50(.33)	.19(.16)	.12(.09)	.38(.34)	.43(.30)	.40(.33)	.34(.21)	
True $\beta^{Z\zeta} = .4$.30(.13)	.42(.17)	.31(.13)	.27(.11)	.42(.19)	.42(.18)	.41(.18)	.41(.16)	
True $\beta^{Z\zeta} = .6$.37(.12)	.41(.13)	.38(.12)	.36(.11)	.41(.13)	.41(.13)	.41(.13)	.41(.13)	
True $\beta^{Z\zeta} = .8$.39(.13)	.40(.13)	.39(.13)	.39(.12)	.40(.13)	.40(.13)	.40(.13)	.40(.13)	

 $J = 100, n_j = 10, \text{ true value between-effects} = .4$ ¹The estimates are averaged over 500 replications ζ should be replaced by the manifest group mode of Z in case of mode aggregation

Table 2

True and Estimated Proportion of Classification Errors

R_{entr}^2	True $\beta^{Z\zeta}$	Modal		Proportional	
		True	Estimated (SD) ¹	True	Estimated (SD) ¹
.24	.20	.27	.21(.08)	.35	.29(.09)
.35	.25	.22	.18(.05)	.30	.25(.06)
.45	.30	.18	.15(.03)	.24	.22(.04)
.55	.35	.14	.13(.03)	.19	.18(.04)
.64	.40	.11	.10(.02)	.15	.15(.03)
.88	.60	.04	.03(.01)	.05	.05(.01)
.97	.80	.01	.01(.00)	.01	.01(.01)

 $J=100, n_j = 100$, true value between-effects = .4¹The estimates are averaged over 500 replications

Table 3

Estimates Between-Association ζ_{1j} and ζ_{2j} ($a_{\zeta_1\zeta_2}$)

$a_{Z_1Z_2}$	1-step		BCH modal		BCH prop		ML modal		ML prop	
	yes	no	yes	no	yes	no	yes	no	yes	no
.0	.40	.40	.39	.40	.40	.40	.39	.40	.40	.40
.2	.40	.44	.40	.43	.40	.44	.40	.43	.40	.44
.4	.40	.50	.40	.48	.40	.50	.40	.48	.40	.50
.6	.41	.60	.42	.56	.43	.59	.42	.56	.43	.59

 $J = 10000, n_j = 100, \text{ true value } a_{\zeta_1\zeta_2} = 0.4$ yes = $a_{Z_1Z_2}$ is incorporated in measurement modelno = $a_{Z_1Z_2}$ is not incorporated in measurement model

Table 4

Estimates Between-Effects Under Different True Values of Indicators ($\beta^{Z_1\zeta_1}$ and $\beta^{Z_2\zeta_1}$) and Within-Association Z_{1ij} and Z_{2ij} ($a_{Z_1Z_2}$) when Within-Association is not Modeled

	1-step		BCH modal		BCH prop		ML modal		ML prop	
$\beta^{Z_1\zeta_1}$ and $\beta^{Z_2\zeta_1}$.6	.2	.6	.2	.6	.2	.6	.2	.6	.2
$a_{Z_1Z_2}$.2	.6	.2	.6	.2	.6	.2	.6	.2	.6
$\beta^{\zeta_1 X}$.41	.34	.41	.35	.41	.35	.41	.35	.41	.35
$\beta^{\zeta_2 X}$.39	.33	.39	.34	.39	.34	.39	.34	.39	.34
$a_{\zeta_1\zeta_2}$.38	.91	.38	.81	.38	.88	.38	.81	.38	.87
$\beta^{Y\zeta_1}$.40	.32	.40	.34	.40	.33	.40	.33	.40	.33
$\beta^{Y\zeta_2}$.40	.30	.40	.32	.40	.31	.40	.32	.40	.31
β^{YX}	.44	.45	.44	.44	.44	.44	.44	.45	.44	.44

$J = 10000$, $n_j = 100$, true value between-effects = .4

Table 5

Class sizes and Class-Specific Response Probabilities Measurement model First Step

	(a) Job control				(b) Job satisfaction				
	Group-level				Group-level				
	latent classes				latent classes				
	1	2	3	Overall	1	2	3	Overall	
Size	.35	.56	.09		Size	.61	.23	.16	
<i>Job control</i>					<i>Job satisfaction</i>				
Low	.70	.13	.01	.32	Low	.88	.38	.34	.69
Medium	.30	.85	.66	.64	Medium	.08	.62	.00	.19
High	.00	.02	.33	.04	High	.04	.00	.66	.12

Table 6

Bias-Adjusted ML Parameters Structural Model

(a)			(b)		
$\beta_l^{\zeta_1 X}$	b	se	$\beta_n^{\zeta_2 X}$	b	se
$\beta_2^{\zeta_1 X}$	0.18	0.29	$\beta_2^{\zeta_2 X}$	0.67	0.34
$\beta_3^{\zeta_1 X}$	1.54	0.57	$\beta_3^{\zeta_2 X}$	-0.79	0.45
$\beta_2^{\zeta_1 X}$	0.98	0.31	$\beta_2^{\zeta_2 X}$	1.31	0.32
$\beta_3^{\zeta_1 X}$	1.96	0.58	$\beta_3^{\zeta_2 X}$	-0.32	0.44
(c)			(d)		
$a_l^{\zeta_1 \zeta_2}$	b	se	β_p^Y	b	se
$a_2^{\zeta_1 \zeta_2}$	-0.79	0.49	$\beta_2^{Y \zeta_1}$	0.26	0.25
$a_2^{\zeta_1 \zeta_2}$	-1.23	0.77	$\beta_2^{Y \zeta_1}$	0.18	0.41
$a_3^{\zeta_1 \zeta_2}$	-0.15	1.04	$\beta_2^{Y \zeta_2}$	1.08	0.28
$a_3^{\zeta_1 \zeta_2}$	1.68	0.87	$\beta_2^{Y \zeta_2}$	-0.20	0.34
			$\beta_2^{Y X}$	0.27	0.15
			$\beta_2^{Y X}$	0.30	0.15

Note: $X = \text{Enriched job design}$, $\zeta_1 = \text{Job control}$,
 $\zeta_2 = \text{Job satisfaction}$, and $Y = \text{Labor productivity}$

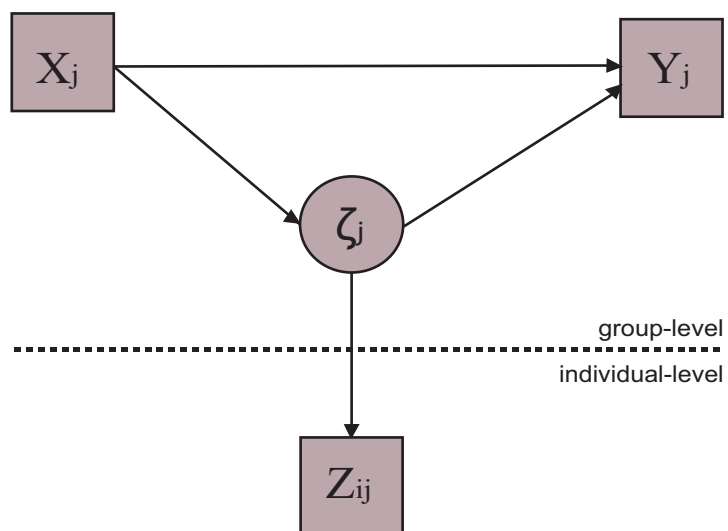


Figure 1. Micro-macro latent variable model with one individual-level predictor.

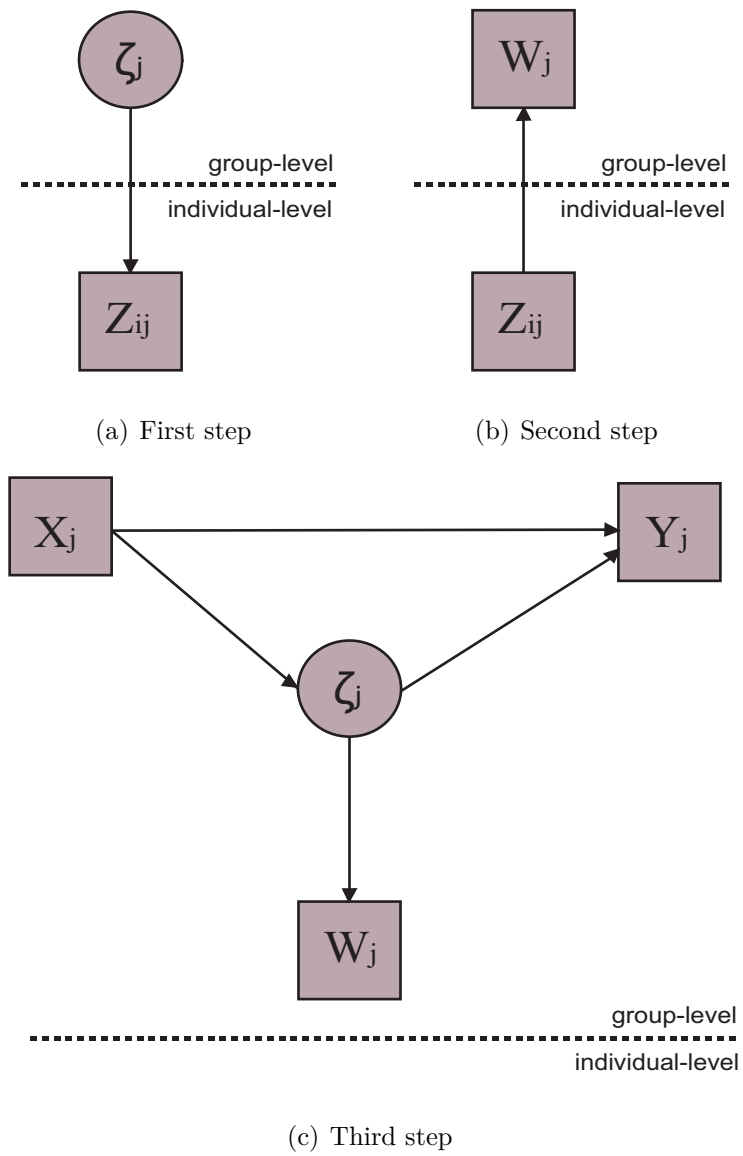


Figure 2. Graphical representation of stepwise procedure.

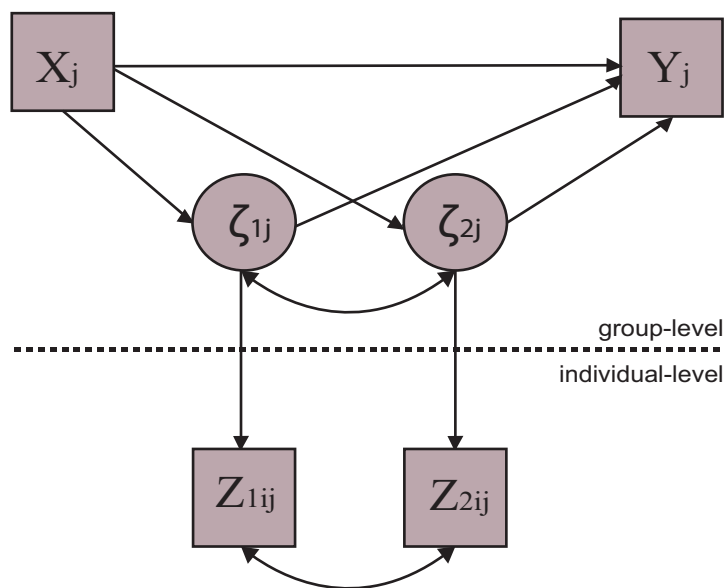


Figure 3. Micro-macro latent variable model with multiple macro-level latent variables.

Appendix A

Technical Details on the Computation of $P(W = t|\zeta = q)$

This Appendix shows how $P(W = t|\zeta = q)$ is computed using the classification information from the second step of the stepwise analysis. A more detailed description is provided by Bakker et al. (2013) and Vermunt (2010).

Let $P(\zeta_j = q|\mathbf{Z}_j)$ denote the posterior class membership probability for group j and $P(W_j = t|\mathbf{Z}_j)$ denote the probability by which a group is assumed to belong to class t of W_j given the applied assignment rule. Using the modal class assignment rule, also called modal a posteriori assignment (MAP), groups are assigned to that category of W_j for which $P(\zeta_j = q|\mathbf{Z}_j)$ is largest:

$$P(W_j = t|\mathbf{Z}_j) = \begin{cases} 1 & \text{if } P(\zeta_j = t|\mathbf{Z}_j) > P(\zeta_j = s|\mathbf{Z}_j) \forall s \neq t \\ 0 & \text{otherwise} \end{cases} . \quad (8)$$

Using the proportional assignment rule, each group is assumed to belong to a particular latent class with a probability equal to the posterior membership probability for the class concerned. Therefore,

$$P(W_j = t|\mathbf{Z}_j) = P(\zeta_j = t|\mathbf{Z}_j) . \quad (9)$$

The probability of being assigned to class t conditional on belonging to the true class q , $P(W = t|\zeta = q)$, is theoretically defined as:

$$P(W = t|\zeta = q) = \frac{\sum_{\mathbf{Z}} P(\mathbf{Z})P(\zeta = q|\mathbf{Z})P(W = t|\mathbf{Z})}{P(\zeta = q)} . \quad (10)$$

Note that the sum is taken over all possible patterns of \mathbf{Z} . Because the number of possible patterns can be very large, it is more practical to take the sum over the data pattern of the groups present in the available data sets, which yields:

$$P(W = t|\zeta = q) = \frac{\sum_{j=1}^J P(\zeta_j = q|\mathbf{Z}_j = z_j)P(W_j = t|\mathbf{Z}_j = z_j)^{\frac{1}{J}}}{P(\zeta_j = q)} . \quad (11)$$

As shown by Vermunt (2010), when the specified model is correct, the theoretical and empirical definition of $P(W = t|\zeta = q)$ provide very similar results.

Appendix B

Latent GOLD 5.0 Syntax Data example

To perform a bias adjusted stepwise analysis on a micro-macro model with two micro-level predictors in Latent GOLD 5.0 (Vermunt & Magidson, 2013), the data need to be structured in a long file format with the number of rows equal to the number of individuals. An identifier variable, here labeled `id`, is needed to identify which individuals belong to which group. The scores on the group-level variables are only assigned to a single group member, for convenience the first group member, so that $Y_{ij} = Y_j$ for $i = 1$, and Y_{ij} is missing for $i \neq 1$, and $X_{ij} = X_j$ for $i = 1$, and X_{ij} is missing for $i \neq 1$.

The relevant parts of the syntax for the first-step measurement model are:

`options`

```
<default settings>
```

```
outfile 'step3data.txt' classification keep y, x;
```

`variables`

```
caseid id;
```

```
dependent z1 nominal, z2 nominal;
```

```
latent zeta1 nominal 3, zeta2 nominal 3;
```

`equations`

```
zeta1 <- 1;
```

```
zeta2 <- 1;
```

```
z1 <- 1 + zeta1;
```

```
z2 <- 1 + zeta2;
```

```
z1 <-> z2;
```

To save the posterior class membership probabilities to a data file, one has to add the command `outfile 'datastep3.txt' classification` to the `options` section. The command `keep` is used to add the variables from the structural part of the model, that are not used in the first-step model, to the output dataset `datastep3.txt` as well. For

the remaining part, one can use the default settings for the `options`.

In the `variables` section, the `id` variable should be defined as the `caseid`. In the same section, a list of the dependent and latent variables should be provided. For nominal latent variables, the number of latent classes is specified after the definition of the scale type.

The regression equations of the first step model are formulated in the `equations` section of the syntax. These include the equations defining the measurement part of the model (`z1 <- 1 + zeta1` and `z2 <- 1 + zeta2`), together with the intercepts for the latent variables (`zeta1 <- 1` and `zeta2 <- 1`), and an equation describing the within-association among the micro-level predictors (`z1 <-> z2`).

The relevant parts of the syntax to estimate the bias corrected third-step structural model are:

```
options
  <default settings>
  step3 ml modal simultaneous;
variables
  dependent y nominal;
  independent x nominal;
  latent zeta1 nominal posterior=(zeta1#1 zeta1#2 zeta1#3),
         zeta2 nominal posterior=(zeta2#1 zeta2#2 zeta2#3);
equations
  zeta1 <- 1 + x;
  zeta2 <- 1 + x;
  zeta1 <-> zeta2;
  y <- 1 + zeta1 + zeta2 + x;
```

This syntax needs to be run on the data file `datastep3.txt`, that was created in the previous step of the analysis. By default, records with missing values are excluded from the analysis, which results in keeping only the first record of each group, and thus

ensures that the analysis is performed at the group-level. The `step3` command specifies the options to be used in the step-three analysis. These concern the correction method (either `none`, `ml` or `bch`) and the assignment rule (`modal` or `prop`). The command `simultaneous` is needed to make sure that all equations from the `equations` section are estimated at once rather than one by one.

In the `variables` section, the dependent, independent and latent variables from the structural part need to be specified. The two latent variables are connected to the stored posterior membership probabilities from the data file using the commands `posterior=(zeta1#1 zeta1#2 zeta1#3)` and `posterior=(zeta2#1 zeta2#2 zeta2#3)`. Finally, all equations from the structural part of the model are specified under `equations`.