

PRO GRADU -THESIS

**Tytti Pasanen**

**Two-Level Structural Equation Modeling with  
Non-Normal Observed Variables for Assessing Poverty  
in Laos**

UNIVERSITY OF TAMPERE  
School of Information Sciences  
Statistics  
September 2012



University of Tampere

School of Information Sciences

PASANEN, TYTTI: Two-level Structural Equation Modeling with Non-Normal Observed Variables for Assessing Poverty in Laos

Master's thesis, 53 p., 9 appendix p.

Statistics

September 2012

---

## Abstract

This thesis introduces a recently developed method, two-level structural equation modeling with non-normal observed variables. The method is applied to household poverty assessment with a survey (n=1564) collected in Lao People's Democratic Republic in 2011. Due to multi-stage clustering in the sampling design, the assumption of the observations' independence is violated. Multilevel modeling ensures this design effect is not ignored.

Structural equation modeling is a flexible method combining confirmatory factor analysis and path analysis, with the ability to model complex causal relations. If the model parameters are estimated with weighted least squares, one is able to model ordinal observed variables often found in survey data. Multilevel methods are currently developed for two-level modeling. They allow examining variation in a data on two levels simultaneously. Multilevel models divide observed variables into two components. Cluster means are modeled on the between level and individual effects, that is, the deviations from the cluster means, are modeled on the within level.

Poverty can be defined in a number of ways. This thesis examines the relation between two poverty indicators, monetary expenditure and the multidimensional poverty index. The latter is modeled by a latent factor structure with three parceled indicators: health, education and living conditions. On the household level, the effect of household size on poverty indicators is considered. On the village level, the number of services and infrastructure are assumed to influence the poverty measures. The final obtained model is valid in terms of the fit indices designed for model assessment. The results show that expenditure and multidimensional poverty are significantly related on household and village levels but the dependencies and relations are stronger on the village level. Infrastructure has a significant impact on the level of poverty in a village. Conclusively, the study suggests that poverty can be decreased effectively by improving communal services.

**Keywords** covariance structure models, latent variable modeling, multilevel modeling, confirmatory factor analysis, multivariate analysis, socio-economic indicators



# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Structural Equation Modeling</b>	<b>9</b>
2.1	Model Specification . . . . .	10
2.2	On Measurement Models . . . . .	11
2.3	Ordinal Observed Variables . . . . .	13
2.4	Weighted Least Square Estimator . . . . .	15
<b>3</b>	<b>Multilevel Structural Equation Modeling</b>	<b>17</b>
3.1	Multilevel Samples . . . . .	18
3.2	Robust Weighted Least Squares Estimation . . . . .	19
3.2.1	Model Specification . . . . .	20
3.2.2	Sample Estimates . . . . .	21
3.2.3	Structural Model Estimation . . . . .	22
<b>4</b>	<b>Model Building</b>	<b>25</b>
4.1	Identification . . . . .	26
4.2	Comparison Models . . . . .	27
4.3	Model Assessment . . . . .	28
<b>5</b>	<b>Defining Poverty</b>	<b>32</b>
5.1	Multidimensional Poverty Index . . . . .	33
5.2	Indicators of MPI . . . . .	34
5.3	Structural Model of Poverty in Laos . . . . .	37
<b>6</b>	<b>Results</b>	<b>40</b>
6.1	Sample . . . . .	41
6.2	Model Comparison . . . . .	42
6.3	Final Model . . . . .	44
6.4	Discussion . . . . .	46
<b>7</b>	<b>Conclusions</b>	<b>49</b>
	<b>Bibliography</b>	<b>50</b>
<b>A</b>	<b>Expectation-Maximisation Algorithm</b>	<b>54</b>
<b>B</b>	<b>Maximum Likelihood EM Algorithm</b>	<b>55</b>

C	Comparison of the indicators in multidimensional poverty index and FFRC data	57
D	Distributions of the measured variables in the MSEM model	58
E	Proof of the identifiability of a between level model using the algebraic solution	60

# 1 Introduction

Poverty is a complex phenomenon which can be defined in multiple ways, each capturing a different dimension of the subject. This thesis proposes a new approach for assessing poverty. The relationship between two poverty definitions is examined by a recently developed method, two-level structural equation modeling with non-normal observed variables, by Asparouhov & Muthén (2007). The method is a combination of multilevel modeling and structural equation modeling, traditionally applied in the fields of psychology and education. This thesis demonstrates how the method is also applicable in social sciences.

Many quantitative studies on poverty have focused on defining relevant livelihood factors (for example, Mtapuri 2011) or exploring two-way relations between these factors, as in the study by Dasgupta, Deichmann, Meisner & Wheeler (2005). Considering the holistic nature of the topic, a method that takes into account several livelihood-related issues at once is preferable in order to receive a better understanding of the nature of poverty.

Rarely used in development studies, structural equation modeling (SEM) is a method that allows complex cause and effect relations in a single model. It is an umbrella term for models that can be constructed as combinations of observed variables, latent factor models and path analysis with varying causal relations. While flexible in theory, the construction of this type of model is often time-consuming, as will be discussed in the following chapters.

Village type determines the availability and importance of certain key aspects related to livelihoods, such as energy and services. Therefore, for reasons out of the villagers' power of decision, some services are unavailable by default. Hence, in this study, the results are examined at both household and village levels simultaneously by incorporating multilevel modeling into traditional SEM. The aim is to explore the relations of the chosen variables, and estimate the proportions of the relations that are explained by the village, and those that are explained by the households themselves.

The data was collected by an extensive nationwide survey in one of the poorest countries in the world, Lao People's Democratic Republic (PDR), in 2011. Households and village heads were interviewed on topics related to livelihoods, energy use and environmental changes. Similar, recent household data from Laos is not currently available from any other source, and the main findings of the survey are yet to be published. Consequently, this thesis provides novel results of poverty in Laos, not only because of the new method but also the newly collected data. With 69 % of the households residing in rural ar-

eas, often with scarce infrastructure, and 44 % of the population considered income poor, Lao PDR is an ideal location for examining the nature of poverty (Government of Laos 2009; Alkire & Santos 2010b).

This thesis begins with an introduction to the statistical background of the applied method. The traditional single-level SEM is presented first in chapter 2. After briefly describing the approach, the chapter's main focus is on the issues that are common with conducting multilevel SEM. Treatment of observed items, whether it be factor analysis, parceling or transforming ordinal variables into their underlying continuous equivalents, is discussed. Model estimation by weighted least squares is also explained.

An overview of multilevel modeling is given in chapter 3, with issues to consider when analyzing clustered data addressed. MSEM estimation procedure for non-normal observed variables is shown in detail at the end of this chapter. The method involves specifying the theoretical model, calculating the sample estimates, and, finally, estimating the structural model parameters. None of the stages are simple or straightforward.

In chapter 4, themes related to model building are discussed. Results are not interpretable if the model is not statistically identified, or the fit of the model is inappropriate. The chapter introduces tools for assuring the content validity of a model.

The concept of poverty is discussed in chapter 5. Model building follows the principles of the multidimensional poverty index (MPI), launched by United Nations Development Programme (UNDP) in 2010. Similarities and differences of the poverty indicators in the MPI, and the data in this thesis are evaluated. The structural model to be tested is proposed at the end.

Chapter 6 exhibits the results. The model specification process is explained in detail. The fit and the parameter estimates of the final model are critically evaluated and interpreted, and the results and their relevance are discussed at the end. Finally, chapter 7 concludes the thesis with summarising the main findings and suggestions.



## 2 Structural Equation Modeling

The fundamental idea behind structural equation modeling (SEM) is to analyse the covariance structure of the selected observed and latent variables, rather than values of individual cases. In an ideal model, the difference between sample covariance and the hypothetical covariance predicted by the model is minimal. The covariance matrix is assumed to consist of functions of a set of parameters. In its simplest form, the hypothesis to be tested is

$$(2.1) \quad \Sigma = \Sigma(\boldsymbol{\theta})$$

where  $\Sigma$  is the population covariance matrix of the observed variables,  $\boldsymbol{\theta}$  is the vector of the model parameters and  $\Sigma(\boldsymbol{\theta})$  is the covariance matrix, written as a function of  $\boldsymbol{\theta}$ . (Bollen 1989). In practice, the population covariance matrix is unknown and the hypothesis testing is based on the sample covariances  $\mathbf{S}$  and the estimated covariance structure  $\Sigma(\hat{\boldsymbol{\theta}})$ .

As indicated by its name, a structural equation model consists of a set of equations, which contain structural parameters and random variables. The structural parameters in the population are estimated based on the known properties, variances and covariances, of the observed variables. Not all parameters are necessarily estimated, for the researcher can choose to add constraints on them.

In regular linear models, the variables are generally regarded as independent, often denoted by  $\mathbf{x}$ , and dependent, denoted by  $\mathbf{y}$ , according to the causal relationships modeled. In structural equation models one variable can represent both cause and effect of others, and the variables in SEM are perceived as *exogenous* and *endogenous*. The distinction between these has been considered fundamentally important, yet debatable (Pearl 2010). Simplifying the issue, exogenous variables are causally independent from the other variables in a particular model, whereas endogenous variables are, either completely or partially, affected by at least one of the other variables in that model.

In this chapter, the general form of a structural equation model is presented first. It is followed by an introduction to different ways of forming the latent constructs. Finally, I present an estimation method for datasets containing ordinal observed variables, a situation encountered often when analysing survey data.

## 2.1 Model Specification

A structural equation model consists of two parts. The *latent structure model* defines the relationships between the latent constructs in the model. The indicators of the latent structures are expressed in the *measurement model*.

The main interest of a researcher is fairly often in the latent structure model. It represents the relations between the theoretical concepts whose intensities are tested. In this section, the general forms of both latent structure and measurement models are presented. In the latter, there are further underlying issues which are discussed in the next section. The contents of this section derive from Bollen's comprehensive manual published in 1989.

The latent model can be written as

$$(2.2) \quad \boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

where  $\boldsymbol{\eta}$  is a vector of endogenous and  $\boldsymbol{\xi}$  a vector of exogenous latent random variables, whose coefficient matrices are  $\mathbf{B}$  and  $\boldsymbol{\Gamma}$ , respectively. That is,  $\mathbf{B}$  and  $\boldsymbol{\Gamma}$  contain the effects their multiplicands have to the latent endogenous variables. For example, the coefficient  $\beta_{ij}$  is the direct impact that  $\eta_j$  has on  $\eta_i$ , given a one unit change in the previous, and everything else held constant. Similarly,  $\gamma_{ji}$  is the effect that  $\xi_i$  has on  $\eta_j$ . The main diagonal of  $\mathbf{B}$  is always zero, by reason of the idea that a concept cannot influence itself immediately without other interventions occurring first. This leads to the assumption that  $(\mathbf{I}-\mathbf{B})$  is a non-singular matrix whose inverse exists.

Other assumptions in 2.2 presume that the expected values of  $\boldsymbol{\eta}$ ,  $\boldsymbol{\xi}$  and  $\boldsymbol{\zeta}$ , the disturbance vector, are 0, and that  $\boldsymbol{\zeta}$  is not correlated with the exogenous variables in  $\boldsymbol{\xi}$ . In the following chapters, the covariance matrices of  $\boldsymbol{\xi}$  and  $\boldsymbol{\zeta}$  are denoted by  $\boldsymbol{\Phi}$  and  $\boldsymbol{\Psi}$ , respectively.

The measurement model connects the latent random variables to their observed indicator variables. The general construct is expressed as

$$(2.3) \quad \boldsymbol{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$$

$$(2.4) \quad \boldsymbol{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

where  $\boldsymbol{x}$  and  $\boldsymbol{y}$  contain the indicator variables of the latent variable vectors  $\boldsymbol{\xi}$  and  $\boldsymbol{\eta}$ , and  $\boldsymbol{\delta}$  and  $\boldsymbol{\epsilon}$  are their measurement errors. Both  $\boldsymbol{\Lambda}$  matrices denote the coefficients that relate the observed variables to the latent vectors. Thus, they represent the expected change in an observed variable after a one unit change in the corresponding latent variable, given that everything else remains constant.

Assumptions in the measurement model are similar to those in the latent model. Expected values of  $\boldsymbol{\eta}$ ,  $\boldsymbol{\xi}$  and the disturbances  $\boldsymbol{\delta}$  and  $\boldsymbol{\epsilon}$  are 0. The disturbance vectors are not allowed to correlate with the latent variables or each other. Covariance matrices of  $\boldsymbol{\delta}$  and  $\boldsymbol{\epsilon}$  are, respectively, denoted with  $\boldsymbol{\Theta}_\epsilon$  and  $\boldsymbol{\Theta}_\delta$ .

Solving  $\boldsymbol{\eta}$  from 2.2, one obtains the equation  $\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})$ . For convenience,  $\mathbf{A}$  is an abbreviation of  $(\mathbf{I} - \mathbf{B})^{-1}$ . Using these formulae, the covariance matrix can be decomposed into the following four elements:

$$\begin{aligned}
\boldsymbol{\Sigma}(\boldsymbol{\theta}) &= \begin{bmatrix} \boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) \end{bmatrix} \\
&= \begin{bmatrix} E(\mathbf{y}\mathbf{y}') & E(\mathbf{y}\mathbf{x}') \\ E(\mathbf{x}\mathbf{y}') & E(\mathbf{x}\mathbf{x}') \end{bmatrix} \\
&= \begin{bmatrix} E\left[(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta}'\boldsymbol{\Lambda}'_y + \boldsymbol{\epsilon}')\right] & E\left[(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\xi}'\boldsymbol{\Lambda}'_x + \boldsymbol{\delta}')\right] \\ E\left[(\boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\eta}'\boldsymbol{\Lambda}'_y + \boldsymbol{\epsilon}')\right] & E\left[(\boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\xi}'\boldsymbol{\Lambda}'_x + \boldsymbol{\delta}')\right] \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Lambda}_y\mathbf{A}\left[\boldsymbol{\Gamma}E(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Gamma}' + E(\boldsymbol{\zeta}\boldsymbol{\zeta}')\right]\mathbf{A}'\boldsymbol{\Lambda}'_y + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') & \boldsymbol{\Lambda}_yE\left[(\mathbf{A}\boldsymbol{\Gamma}\boldsymbol{\xi})\boldsymbol{\xi}'\right]\boldsymbol{\Lambda}'_x \\ \boldsymbol{\Lambda}_xE\left[\boldsymbol{\xi}(\boldsymbol{\xi}'\boldsymbol{\Gamma}'\mathbf{A}')\right]\boldsymbol{\Lambda}'_y & \boldsymbol{\Lambda}_xE(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Lambda}'_x + E(\boldsymbol{\delta}\boldsymbol{\delta}') \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Lambda}_y\mathbf{A}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{A}'\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_\epsilon & \boldsymbol{\Lambda}_y\mathbf{A}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{A}'\boldsymbol{\Lambda}'_y & \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x + \boldsymbol{\Theta}_\delta \end{bmatrix}.
\end{aligned}$$

## 2.2 On Measurement Models

Structural equation models cover a variety of model specifications. While the latent model should be the foundation of the research, the underlying measurement model may include various types of observed constructs. In this section, I introduce three types of measurement models. First is the most traditional, and least controversial, confirmatory factor analysis (CFA). It is followed by single-indicator latent factors. Third, I discuss the issue of using parceled items, either as single factors or as part of a factor construct.

**Factor Constructs** Often in structural equation models, each latent variable is assigned multiple indicators. The procedure is better known as confirmatory factor analysis. In CFA, the researcher determines beforehand the indicator variables of the latent construct. Each observed item has its unique loading factor and a residual parameter. Therefore, when modeling latent relations, it is possible to detect errors that occur in the measurement of the constructs, rather than in the model itself.

Confirmatory factor analysis also allows flexibility on the specification of the model. For example, imposing constraints such as correlating errors of measurement are possible with CFA. A common restriction is to force one of the loadings of a latent construct to equal to 1. This type of constraint may not only help to identify the model but it also ensures that all indicators are measured on the same scale. (Bollen 1989).

**Single-Indicators** A latent construct may also be indicated by a single indicator. What is problematic in this approach is that the observed values are now perceived as exact measures of the latent concept. No error on the indicator

level is modeled, even though most measurable concepts are, in reality, imperfectly measured and recorded. If the reliability of a construct is known from previous research or if it can be estimated, the uncertainty related to the use of a single indicator can be taken into account. Unfortunately, the reliability or the means to estimate the reliability of a measure are rarely available. (Bollen 1989).

Possible reasons for a researcher to proceed the model building with single indicators relate to practical issues such as data availability (Bollen 1989). If the reliability remains unknown, it makes no difference to the model estimation whether a single-indicator latent variable or an observed manifest variable is included. For identification purposes, the latent factor loading and the error of the latent variable need constraining nevertheless, and the number of estimated parameters remains the same. (Little, Cunningham, Shahar & Widaman 2002).

**Parceling** Parceled item is defined as an aggregated indicator that comprises of the sum or average of more than one observed items. It is a technique commonly used, but less often critically viewed in the applied literature. A publication by Little et al. (2002) discusses the advantages and risks of modeling with parceled constructs.

When modeling with parcels rather than individual variables, many indices of model fit are expected to become more acceptable. Little et al. (2002) have summarized three reasons why this happens. Firstly, the models using parceled variables have fewer estimated parameters. They are more parsimonious, whereas many indices add penalty for additional parameters. Secondly, the chances of encountering dual loadings (that is, an observed variable loading on two or more latents) or correlation of the residuals are fewer. Finally, parceling reduces the sources of sampling error. What is more, parcels can be used to obtain a just-identified, thus, undisputedly unique, solutions instead of an over-identified factor that may have many alternative solutions.

Parcels can be constructed from uni- or multi-dimensional observed items. Of the two, the parcels that define a multidimensional construct have been found more dubious and defining such should be conducted with caution. As an indicator of a multidimensional construct, the parcel itself is likely to be a sum of more than one substantive constructs. Thus, it may provide biased loading estimates. The latent construct might also have fewer specified dimensions than, in reality, underlie it.

In essence, whether a parcel construct is appropriate depends on the approach to the latent variable modeling. If the main focus is on the relations among the items, parceling is not recommended. Ignoring correlated residuals or dual loadings, the patterns of the observed data cannot be fully understood. Instead, if the interest is mainly in the relations between the latent constructs, parceling is more willingly warranted. From this viewpoint, observed variables are but tools that enable constructing a measurement model for the desired latent structure. Each construct is content-specific and the chosen measurement,

in each construct, should be justified separately.

Parceling may relate to multiple and single-indicator factor modeling in various ways. It serves as an alternative to confirmatory factor analysis if the factor indicators are aggregated into one parceled variable, which can be used as either a single-indicator latent or a manifest variable. Parcels may complement and reduce the complexity of CFA by reducing the number of indicators (Yang, Nay & Hoyle 2010). Parceling may also prevent having to construct a higher order factor model. If the highest order latent factor is multidimensional and each dimension is a latent construct itself, the model might remain easier to interpret and converge if the lower level factors are parceled items, instead of latent factors. Yang et al. (2010) found that parceling is most effective when the indicators consist of five response categories and the parcels are used as indicators of latent constructs.

## 2.3 Ordinal Observed Variables

The assumption of multivariate normality of the observed variables is not always met in reality. Even though the distributions in theory would be normal, there might be errors in the measurement or the sampling that disturb the observed distribution. If the level of measurement is dichotomous or ordinal, the assumption of normality does not hold by default. In such case, the parameters in the model cannot be estimated until the observed variables have been adjusted.

In a method designed to manage non-continuous observed variables, the observed items are regarded as indicators of their underlying continuous latent variables, denoted by '\*' (e.g.  $\mathbf{x}^*$ ). The equation  $\Sigma = \Sigma(\theta)$  cannot hold anymore, instead, the main hypothesis needs to be expressed as

$$\Sigma^* = \Sigma(\theta)$$

where  $\Sigma^*$  is the population covariance structure of the continuous latent variables. (Bollen 1989).

When modeling dichotomous or ordinal observed variables, the formulae expressing the measurement model, 2.3 and 2.4, do not necessarily hold. The distribution of an ordinal  $\mathbf{x}_q$  or  $\mathbf{y}_p$  might differ substantially from the distribution of the underlying, continuous  $\mathbf{x}_q^*$  or  $\mathbf{y}_p^*$ . (Bollen 1989). For example, categories in ordinal variables might not be of equal length which leads to a violation of the original assumption. If the underlying continuous variables are assumed multinormal, replacing  $\mathbf{x}$  with  $\mathbf{x}^*$  in 2.3 and  $\mathbf{y}$  with  $\mathbf{y}^*$  in 2.4 corrects the equations (Bollen 1989).

Consequently, ordinal observed variables cannot be in linear relations with their underlying latent continuous variables. A nonlinear function needs to be fitted to delineate the relationship (Bollen 1989). As a result, the variances and covariances of the observed and latent variables are not in a linear relation either, and the evaluation of their true scales becomes complex. The general

way to solve this problem is to standardise the latent, continuous variables to the mean of 0 and variance of 1. Consequently, the covariance estimates and the correlations become equal.

If two variables are both ordinal by origin, the correlation of their continuous equivalents is called *polychoric* correlation. A special case of this correlation is *tetrachoric* correlation, referring to two dichotomous variables. If one observed variable is ordinal and the other continuous, correlation of their latent continuous variables is called *polyserial* correlation. (Bollen 1989).

All of these can be estimated with one- or two-stage procedures, the latter being more computationally efficient, and, hence, often the only estimation method offered by SEM softwares. The estimates from the two methods tend to be very close, although, statistically, the estimates from the two-stage method are not asymptotically efficient. Nevertheless, they are consistent and asymptotically normal. (Maydeu-Olivares, García-Forero, Gallardo-Pujol & Renom 2009).

**Two-Stage Estimation Method** The two-stage method was first presented by Olsson (1979), and it includes assigning a monotonic function to estimate the values of the underlying continuous variables, based on the corresponding ordinal discrete variables. At the beginning, thresholds are estimated for each marginal distribution of standardised  $\mathbf{x}_q^*$  and  $\mathbf{y}_p^*$  ( $q = 1, 2, \dots, Q$  and  $p=1, 2, \dots, P$ , where  $Q$  and  $P$  are the numbers of exogenous and endogenous observed, non-continuous variables in  $\mathbf{x}$  and  $\mathbf{y}$ ). For the model to apply, one needs to define such thresholds  $a_k$  for each  $p$  variables for which the following formula holds:

$$(2.5) \quad y_{pi} = k \Leftrightarrow a_{pk-1} < y_{pi}^* < a_{pk}.$$

In 2.5,  $i$  is a subscript of the individuals,  $k=1, 2, \dots, c$ , and  $c$  is the number of categories in  $\mathbf{y}_p$ . Similarly, the number of categories in  $\mathbf{x}_q$  is  $d$ . For simplicity, in the rest of this chapter,  $a_{pk}$  is abbreviated to  $a_k$ . The lowermost and uppermost thresholds are  $a_0=-\infty$  and  $a_c = \infty$ , and the remaining  $c - 1$  thresholds are estimated by the following formula

$$(2.6) \quad a_k = \Phi^{-1} \left( \sum_{i=1}^k \frac{N_i}{N} \right)$$

where  $\Phi$  is the standardized normal distribution function and  $N_i$  the number of cases in the  $i$ th category. (Bollen 1989; Asparouhov & Muthén 2007). Simply put, the thresholds are equivalent to the values of standard normal distribution at the cumulative percentage points of the observed variables.

Jöreskog (1990) states that the aim of the second phase is to estimate the polychoric correlation with maximum likelihood estimation, given the thresholds defined in 2.6. Let  $\pi_{kl}$  denote the probability that an observation falls into cell  $(k, l)$ , which, in the case of bivariate standard normal distribution with correlation coefficient  $\rho$ , is defined as

$$\begin{aligned}\pi_{kl} &= Pr[x_q^* = k, y_p^* = l] = \int_{a_{k-1}}^{a_k} \int_{b_{l-1}}^{b_l} \phi_2(u, v) dv du \\ &= \Phi_2(a_k, b_l) - \Phi_2(a_{k-1}, b_l) - \Phi_2(a_k, b_{l-1}) + \Phi_2(a_{k-1}, b_{l-1})\end{aligned}$$

where  $\phi_2$  is the bivariate normal probability function and  $\Phi_2$  the corresponding distribution function, and  $a_k$  and  $b_l$  denote the thresholds for the variables  $\mathbf{x}_q^*$  and  $\mathbf{y}_p^*$ . From here, the likelihood function to be maximised is in the form of

$$L = C \prod_k^d \prod_l^c \pi_{kl}^{N_{kl}}$$

where  $C$ , in this case, is an irrelevant constant and can be dropped from the formula at this point, and  $N_{kl}$  represents the number of cases in the cell  $(k, l)$ . Taking logarithms on both sides, one obtains the more easily solvable

$$\ln L = \sum_k^d \sum_l^c N_{kl} \ln(\pi_{kl})$$

for which the maximum can be found, as a common procedure, by taking derivatives in terms of the unknown parameter  $\rho$  and solving the zero point. (Olsson 1979). Finally, the estimation becomes a question of solving the equation

$$\begin{aligned}\frac{\partial l}{\partial \rho} &= \sum_{k=1}^d \sum_{l=1}^c \frac{N_{kl}}{\pi_{kl}} \left( \frac{\partial \pi_{kl}}{\partial \rho} \right) \\ &= \sum_{k=1}^d \sum_{l=1}^c \frac{N_{kl}}{\pi_{kl}} [\phi_2(a_k, b_l) - \phi_2(a_{k-1}, b_l) - \phi_2(a_k, b_{l-1}) + \phi_2(a_{k-1}, b_{l-1})] = 0.\end{aligned}$$

Olsson (1979) concluded that the absolute differences between the true correlations and the estimated ones were low in practice. In fact, theoretically, the polychoric correlation matrix is a consistent estimator of  $\Sigma^*$ . This enables testing the hypothesis  $\Sigma^* = \Sigma(\boldsymbol{\theta})$  with the method described in this section (Bollen 1989.)

## 2.4 Weighted Least Square Estimator

In practice, the researcher does not know the parameters in the model, nor the population covariances or variances. The assumed population values are derived from the sample covariance (or correlation) matrix  $\mathbf{S} = \hat{\Sigma}$ . The unknown structural parameters in  $\boldsymbol{\theta}$  are estimated by minimising the differences between the structural covariance matrix  $\Sigma(\boldsymbol{\theta})$  and  $\mathbf{S}$  with an appropriate fitting function  $F$ . (Bollen 1989).

There are many available estimation functions  $F(\mathbf{S}, \Sigma(\boldsymbol{\theta}))$ . Maximum likelihood (ML) is recommended if the observed variables are normally distributed.

Generalized least squares (GLS) and unweighted least squares (ULS) are also suitable alternatives which require less information but whose properties are yet known. If the observed variables in the model are not continuous and normally distributed, the recommended fit function is weighted least squares (WLS). Lei (2009) found WLS to result in relatively small bias unless the sample size is less than 100. If the sample correlation matrix is used in the estimation, WLS and the varieties derived from WLS are the only viable options (Olsson 1979; Bollen 1989; Yu 2002; Hox, Maas & Brinkhuis 2010; Boulton 2011).

For the WLS fitting function,  $\hat{\boldsymbol{\rho}}$  is defined as a vector which contains all non-duplicated elements of the sample covariance or correlation matrix. The vector  $\boldsymbol{\sigma}(\boldsymbol{\theta})$  contains the non-duplicated elements of  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ . Now, the WLS fitting function is

$$(2.7) \quad F_{\text{WLS}} = [\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\boldsymbol{\theta})]' \mathbf{W} [\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\boldsymbol{\theta})].$$

Values of  $\boldsymbol{\theta}$  are chosen in a way that minimises the weighted sum of the deviations between  $\hat{\boldsymbol{\rho}}$  and  $\boldsymbol{\sigma}(\boldsymbol{\theta})$ . This leads to  $\hat{\boldsymbol{\theta}}$  becoming a consistent estimator of  $\boldsymbol{\theta}$ , given that the original assumption  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$  is true.

In order to determine the weight matrix  $\mathbf{W}$ , it is necessary to introduce a matrix  $\mathbf{G}$  which is the asymptotic covariance (or correlation) matrix of  $\hat{\boldsymbol{\rho}}$ , or its consistent estimator. It has been proven that if  $\mathbf{W} = \mathbf{G}^{-1}$ , then  $\hat{\boldsymbol{\theta}}$  obtained by  $F_{\text{WLS}}$  is asymptotically efficient in the class of functions presented in 2.7. (Bollen 1989). If identity matrix is chosen instead ( $\mathbf{W} = \mathbf{I}$ ), one receives the ULS estimator. If only the diagonal elements of  $\mathbf{G}^{-1}$  are chosen and the off-diagonal elements of the weight matrix are 0 ( $\mathbf{W} = \mathbf{G}_0^{-1}$ ), the estimator is diagonal weighted least squares (DWLS). (Asparouhov & Muthén 2007.)

While computationally more efficient than WLS, the asymptotic properties of  $F_{\text{DWLS}}$  differ from the properties of  $F_{\text{WLS}}$ . Consequently, the same assumptions behind the model fit assessment that apply to  $F_{\text{WLS}}$  do not hold with  $F_{\text{DWLS}}$ . Different adjustment methods to correct the bias have been suggested. (Asparouhov & Muthén 2007). The topic will be discussed more in the following two chapters.

The WLS weight matrix  $\mathbf{G}^{-1}$  needs to be positive-definite. Its calculation can become complicated, especially if the number of variables is large. It also requires a rather large N to be consistent. (Jöreskog 1990). The formulation of  $\mathbf{G}$  involves several stages and it has been defined by Olsson (1979), after which the estimated matrix was proven to be applicable for the case of polychoric correlation by Muthén (1984). WLS and polychoric correlations have been found to provide a better fit in the presence of non-normality, for example when analysing ordinal data (Ory & Mokhtarian 2010; Şimşek & Noyan 2012). With a complicated model, computing  $\mathbf{G}$  might prove impossible due to the limitations mentioned above. In such a case, DWLS estimation with  $\mathbf{G}_0^{-1}$  is a feasible alternative.



### 3 Multilevel Structural Equation Modeling

In social sciences, datasets are often structured hierarchically. One can examine individuals nested within different groups and societies they belong to. Even though the main interest of a study is focused on the individual, survey samples are often clustered for time- and cost-effective reasons. This may lead to a violation of the observations' independence of each other, which should not be neglected. One way to avoid this violation is to conduct the analysis on several levels. Multilevel modeling takes into account the hierarchical structure of a given dataset, and its applications cover a number of statistical methods. The two main branches are regression and covariance structure models (Hox 2002). In this chapter, I concentrate on the latter, more precisely, the multilevel structural equation model (MSEM) approach.

There is no single established name for multilevel models. Depending on the field of science, they can be referred to as hierarchical models, multilevel linear models, mixed- or random effects models, random coefficient models or covariance components models. What is common to all of them is that each level forms its own submodel. The submodels express how the variables are connected within the level and across the other levels. Most methods are currently developed for two levels only. (Raudenbush & Bryk 2002.) Originally, multi-level modeling was restricted to the assumptions of linearity and normality. Recent developments, such as the computationally efficient model presented in this chapter, are less restrictive in their assumptions.

Following the guide to multilevel modeling by Hox (2002), a multilevel covariance structure is expressed, in its simplest form, as follows. The data to be structured is denoted by a  $p$ -variate matrix  $\mathbf{y}_{ij}$ , where  $i$  and  $j$ , respectively, indicate individuals and clusters. The observed values, *total score* ( $t$ ), of  $\mathbf{y}_{ij}$  are assumed to consist of two parts:

$$\begin{aligned}
 (3.1) \quad \mathbf{y}_t &= \mathbf{y}_w + \mathbf{y}_b \\
 &= \Lambda_w \boldsymbol{\eta}_{wij} + \boldsymbol{\epsilon}_{wij} + \Lambda_b \boldsymbol{\eta}_{bj} + \boldsymbol{\epsilon}_{bj}
 \end{aligned}$$

where  $\mathbf{y}_b$  is the between groups component, defined as  $\bar{\mathbf{y}}_j$ , and  $\mathbf{y}_w$  is the within group component,  $\mathbf{y}_{ij} - \bar{\mathbf{y}}_j$ . A further decomposition is provided at section 3.2.

There is one major difference in the denotations of an MSEM model compared to the SEM model described in the previous chapter. In a general SEM, the latent variables are divided into endogenous, indicated by the observed

variables  $\mathbf{y}$ , and exogenous, indicated by the observed variables  $\mathbf{x}$ . In MSEM, the main division is if the observed variables exist in one level only (denoted by  $\mathbf{x}$ ) or if they are featured on different levels ( $\mathbf{y}$ ).

The expressions 'random effects model' and 'random coefficient model' give some insight into these multilevel properties of  $\mathbf{y}$ . The division of a variable  $\mathbf{y}_p$  into a cluster level ( $\mathbf{y}_{bp}$ ) and an individual ( $\mathbf{y}_{wp}$ ) effect creates two new independent latent variables. The between group component, that is, the cluster effect, is a disaggregated variable which contains the cluster averages of each group. These cluster averages are regarded as random samples from a population of group intercepts. On the within level, they are specified as random intercepts whereas on the between level, they are random variables modeled separately on that level. This multilevel approach enables examining two types of dependencies simultaneously: relations between different group characteristics and the relations between the individuals, given the cluster effect.

Proceeding from 3.1, the population covariance matrix in the multilevel case also consists of two separate parts:

$$(3.2) \quad \Sigma_t = \Sigma_w + \Sigma_b.$$

In 3.2,  $\Sigma_b$  is the population covariance matrix of the group means in  $\mathbf{y}_b$ , and  $\Sigma_w$  is the covariance matrix of the individual deviations from the group means,  $\mathbf{y}_w$ . (Hox 2002). The approach is also known as *disaggregated modeling*. Ignoring the multilevel structure of the data, one examines solely  $\Sigma_t$  by a conventional SEM or *aggregated modeling*. (Muthén 1997.)

The estimates of the disaggregated population covariances  $\Sigma_w$  and  $\Sigma_b$  cannot be obtained by directly calculating the sample covariances,  $\mathbf{S}_b$  and  $\mathbf{S}_w$ , especially in the case of non-normal observed variables. (Hox 2002). The procedure for estimating  $\Sigma_b$  and  $\Sigma_w$  is provided later in this chapter. Before a more detailed description of the MSEM procedure, I present some common issues that require attention when dealing with clustered data.

### 3.1 Multilevel Samples

Whether a multilevel analysis is reasonable or not, depends mainly on the sample design and the properties of the observed variables. In this section, I present three measures which help to evaluate whether a multilevel model is appropriate. *Intra-class correlation* (ICC) is a measure related to the observed variables, *design effect* concerns the accuracy of the model estimates and *effective sample size* estimates if the design effect is adequately taken into account.

Intra-class correlation describes the share of the variation explained by the covariance between the clusters, compared to the total variance. Thus, the ICC of  $\mathbf{y}_p$ , denoted by  $\rho_p$ , is defined as

$$\rho_p = \frac{\sigma_{bp}^2}{\sigma_{bp}^2 + \sigma_{wp}^2}.$$

If the ICC of a variable is low, specifying the model on more than one levels might be unnecessary. Some studies have suggested 0.05 as the minimum acceptable ICC in this regard. (Hox 2002).

In multilevel sampling designs, the effect of cluster sampling on the variances of the estimates is called *design effect*. Snijders & Bosker (1999) define design effect as the factor by which the total sample size  $n$  needs to be increased in order to achieve the same estimation variance as an equal-sized simple random sample would have. A large design effect implies a relatively large variance for the estimates and is, thus, undesirable. In a two-level design, the formula is

$$(3.3) \quad \text{design effect} = 1 + (n_{clus} - 1)\rho.$$

In 3.3, the cluster size  $n_{clus}$  is assumed to be constant across groups, although, the formula provides decent approximations even if the cluster sizes are variable but not too widely different (Snijders 2005).

Following from 3.3, the sample can become less attractive for two reasons. First, if the ICC increases, that is, the clusters become more homogenous. Second, the average cluster size  $n_{clus}$  increases. (Snijders & Bosker 1999.) If the design effect is considerable but ignored in the analysis, the standard errors of the model estimates are too small, which may lead to false significant results.

*Effective sample size* is a measure which estimates whether the design effect has been taken into account sufficiently in the sampling. It is an estimate of the minimum sample size needed to obtain correct significant results. Efficient sample size ( $n_{eff}$ ) depends on the population's intra-class correlation  $\rho$ , total sample size  $n$  and the average cluster size  $n_{clus}$ , as follows:

$$n_{eff} = \frac{n}{1 + (n_{clus} - 1)\rho}.$$

If  $n_{eff}$  is greater than or equal to the total sample size  $n$ , the design effect has been considered in the sampling sufficiently. (Hox 2002).

Finally, Snijders (2005) and Maas & Hox (2005) argue that the sample size at the higher level is the main limiting characteristic of the design. The argument of the latter is based on a simulation study on the effective sample sizes of multilevel designs. The samples were simulated with varying  $ns$  and ICCs, with each design involving a considerable design effect and ICC. The results showed that if the higher level sample size is more than 50, the estimated coefficients, variance components and standard errors were accurate and unbiased.

## 3.2 Robust Weighted Least Squares Estimation

As argued in 2.4, the weighted least squares estimator is the most appropriate in the case of non-normal data. Its use has been acknowledged and supported by traditional SEM users and software. As for multilevel methods, on the contrary, a WLS estimation technique is fairly recently developed. It is currently only available in the software package Mplus. In this section, I present the theory

and the procedure of this new robust limited-information method. The main findings are based on the work of Asparouhov & Muthén (2007).

The estimation is principally conducted in the same stages as the traditional SEM models. It begins with specifying the structural model to be tested. Second, the sample estimates of the data are calculated. In the case of non-normal, hierarchically structured observed variables, the calculation involves several procedures, including maximum likelihood methods along with the expectation-maximisation (EM) algorithm and numerical integration. The univariate parameters are estimated first, after which bivariate methods are applied. Third, the parameters of the structural model are estimated via minimising the WLS fit function. This final estimation is preceded by expressing the structural model as nested within the model in the second stage.

The procedure is based on the method by Muthén (1984), which has been extended to the two-level case recently. One of its main characteristics is that it splits the estimated model into multiple simple models, where no higher than one- and two-dimensional integration is needed. This way the estimates are obtained computationally efficiently, without compromising on precision, as was concluded by Asparouhov & Muthén (2007). A further advantage of the method is that there are no restrictions on the measurement of the observed variables. Any combination of binary, censored, ordered polytomous and continuous observed variables can be included in the model to indicate the latent structures.

### 3.2.1 Model Specification

In this section, I define the decomposition of the structural model. It begins with specifying the underlying latent variables  $\mathbf{y}^*$ . If the observed  $\mathbf{y}_p$  is normally distributed,  $\mathbf{y}_p = \mathbf{y}_p^*$  holds. The theoretical thresholds are formulated on the same grounds as described in 2.3:

$$(3.4) \quad y_{pij} = k \Leftrightarrow \tau_{pk-1} < y_{pij}^* < \tau_{pk}.$$

The underlying normally distributed latent  $\mathbf{y}_p^*$  is itself a combination of two normally distributed independent latent constructs:

$$(3.5) \quad y_{pij}^* = y_{wpij} + y_{bpij}$$

where  $j = 1, \dots, C$  represent the clusters,  $i = 1, \dots, N_j$  the individuals in each cluster and  $p = 1, \dots, P$  the observed variables that exist on both levels. The individual effect is denoted by  $y_{wpij}$  and the cluster effect by  $y_{bpij}$ , both independent and normally distributed latent variables.

For the structural model, one defines the latent normally distributed vector variables  $\boldsymbol{\eta}_{wij} = (\eta_{w1ij}, \dots, \eta_{wM_1ij})$  on the individual and  $\boldsymbol{\eta}_{bj} = (\eta_{b1j}, \dots, \eta_{bM_2j})$  on the cluster level, with  $M_1$  denoting the number of the within level and  $M_2$  denoting the number of the between level latent constructs. The independent variables,  $Q_1$ -variate  $\mathbf{x}_{wij}$  and  $Q_2$ -variate  $\mathbf{x}_{bj}$ , are expressed as  $x_{wq_1ij}$  and  $x_{bq_2j}$  where  $q_1 = 1, \dots, Q_1$  and  $q_2 = 1, \dots, Q_2$ .

Thus, the individual and the between level structural models are as follows:

$$(3.6) \quad \begin{cases} \mathbf{y}_{wij} &= \mathbf{\Lambda}_w \boldsymbol{\eta}_{wij} + \boldsymbol{\varepsilon}_{wij} \\ \boldsymbol{\eta}_{wij} &= \mathbf{B}_w \boldsymbol{\eta}_{wij} + \mathbf{\Gamma}_w \mathbf{x}_{wij} + \boldsymbol{\xi}_{wij} \end{cases}$$

$$(3.7) \quad \begin{cases} \mathbf{y}_{bj} &= \boldsymbol{\nu}_b + \mathbf{\Lambda}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\varepsilon}_{bj} \\ \boldsymbol{\eta}_{bj} &= \boldsymbol{\alpha}_b + \mathbf{B}_b \boldsymbol{\eta}_{bj} + \mathbf{\Gamma}_b \mathbf{x}_{bj} + \boldsymbol{\xi}_{bj} \end{cases}$$

where the matrices  $\mathbf{\Lambda}_w$ ,  $\mathbf{B}_w$ ,  $\mathbf{\Gamma}_w$ ,  $\boldsymbol{\nu}_b$ ,  $\mathbf{\Lambda}_b$ ,  $\boldsymbol{\alpha}_b$ ,  $\mathbf{B}_b$  and  $\mathbf{\Gamma}_b$  contain the parameters to be estimated. The between level intercept vectors  $\boldsymbol{\nu}_b$  and  $\boldsymbol{\alpha}_b$  are not substantially interesting. The residual matrices  $\boldsymbol{\varepsilon}_{wij}$ ,  $\boldsymbol{\xi}_{wij}$ ,  $\boldsymbol{\varepsilon}_{bj}$  and  $\boldsymbol{\xi}_{bj}$  are independent and normally distributed with zero mean. Their respective covariances are denoted by  $\boldsymbol{\Theta}_w$ ,  $\boldsymbol{\Psi}_w$ ,  $\boldsymbol{\Theta}_b$  and  $\boldsymbol{\Psi}_b$ . In order to obtain identification, the variance of  $\boldsymbol{\varepsilon}_{wpj}$  is fixed at 1 if the  $p$ th variable is categorical. Some other restrictions may also need to be posed to ensure identifiability, depending on the model.

### 3.2.2 Sample Estimates

The calculation of the sample estimates is more complex when the data is hierarchically structured and the observed variables are non-continuous. In this section, I present the formation of the data and a summary of the computational methods required for the calculations.

The data is expressed as a saturated model. It contains no latent constructs  $\boldsymbol{\eta}_{wij}$  or  $\boldsymbol{\eta}_{bj}$ , and full covariance matrices are fitted for the within and the between level variables. The topic of saturation and other reference models will be further discussed in section 4.3.

The thresholds of the categorical variables have the same construction as in the one-level structural equation models in formula 2.5 and the underlying latent  $\mathbf{y}_p^*$  is defined as in the structural model specified in 3.5:

$$(3.8) \quad y_{pij} = k \Leftrightarrow a_{pk-1} < y_{pij}^* < a_{pk}$$

$$(3.9) \quad y_{pij}^* = y_{wpj} + y_{bpj}.$$

As previously described, the saturated model is in the form of

$$(3.10) \quad \begin{aligned} \mathbf{y}_{wij} &= \mathbf{\Pi}_w \mathbf{x}_{wij} + \boldsymbol{\epsilon}_{wij} \\ \mathbf{y}_{bj} &= \boldsymbol{\mu}_b + \mathbf{\Pi}_b \mathbf{x}_{bj} + \boldsymbol{\epsilon}_{bj}. \end{aligned}$$

The residual vectors  $\boldsymbol{\epsilon}_{wij}$  and  $\boldsymbol{\epsilon}_{bj}$  are assumed normally distributed with 0 means and covariance matrices  $\boldsymbol{\Sigma}_w$  and  $\boldsymbol{\Sigma}_b$ , respectively. For identification purposes, if the  $p$ -th variable is categorical, the variance of  $\boldsymbol{\epsilon}_{wpj}$  is fixed at 1 and the mean parameter  $\mu_{pb}$  is fixed at 0.

The parameters in 3.10 are estimated in two stages. The first stage is for the estimation of the univariate parameters: the between means  $\mu_{bp}$ , the thresholds  $a_{pk}$ , the coefficients  $\Pi_{wpq}$  and  $\Pi_{bpq}$ , and the diagonal elements of the residual covariance matrix,  $\Sigma_{wpp}$  and  $\Sigma_{bpp}$ . At the second stage, the covariances of

residuals, that is, the off-diagonal elements of  $\Sigma_w$  and  $\Sigma_b$  are estimated with bivariate likelihood methods, given the univariate estimates.

The method for univariate and bivariate maximum likelihood estimation is generalised from the growth mixture model procedure presented by Asparouhov & Muthén (2008). The method is based on the EM algorithm where the latent variables  $\mathbf{y}_{wij}$  and  $\mathbf{y}_{bj}$  are treated as missing data. The EM algorithm is an iterative method for maximum likelihood estimation when the data are incomplete or complex. Such is the case with hierarchical data. The algorithm involves two steps - expectation and maximisation - which will be continued until acceptable convergence has been obtained. One of the main benefits of using the EM algorithm is that it reduces, either artificially or by forcing, the complexity of maximum likelihood estimation. (McLachlan & Krishnan 1997.) The principal iterative method behind the EM algorithm is shortly described in appendices A and B.

After the two-stage iterative estimation, the univariate and bivariate estimates are combined in a vector  $\mathbf{s}$ . The asymptotic covariance matrix of  $\mathbf{s}$ , denoted by  $\mathbf{G}$ , is computed. The formulation of  $\mathbf{G}$  as a consistent covariance matrix of the estimated parameters has been found and proven by Muthén and Satorra (1995). The covariances are based on the first derivatives of the likelihoods in the univariate and bivariate estimation, given that the bivariate estimates are conditional on the univariate estimates. The inverse of  $\mathbf{G}$  will be used as the weight matrix in the WLS estimation in the final stage.

### 3.2.3 Structural Model Estimation

In the third stage of the two-level WLS estimation, the structural or restricted model parameters are estimated as in Muthén (1984), but with extending the method into two-level-structure. Let us define the model first. The threshold estimation and multilevel structure are expressed as the saturated model 3.8 - 3.10:

$$y_{pij} = k \Leftrightarrow a_{pk-1}^* < y_{pij}^* < a_{pk}^* \quad (3.11)$$

$$\mathbf{y}_{wij} = \mathbf{\Pi}_w^* \mathbf{x}_{wij} + \boldsymbol{\epsilon}_{wij} \quad (3.12)$$

$$\mathbf{y}_{bj} = \boldsymbol{\mu}_b^* + \mathbf{\Pi}_b^* \mathbf{x}_{bj} + \boldsymbol{\epsilon}_{bj}.$$

The structural model is nested within the saturated or unrestricted model. The index '\*\*' refers to a further decomposition which includes the parameters to be estimated. The decomposition is presented next. The merged forms of the theoretical structural model in 3.6 and 3.7 are

$$\mathbf{y}_{wij} = \mathbf{\Lambda}_w (\mathbf{I} - \mathbf{B}_w)^{-1} (\mathbf{\Gamma}_w \mathbf{x}_{wij} + \boldsymbol{\xi}_{wij}) + \boldsymbol{\epsilon}_{wij} \quad (3.13)$$

$$\mathbf{y}_{bj} = \boldsymbol{\nu}_b + \mathbf{\Lambda}_b (\mathbf{I} - \mathbf{B}_b)^{-1} (\boldsymbol{\alpha}_b + \mathbf{\Gamma}_b \mathbf{x}_{bj} + \boldsymbol{\xi}_{bj}) + \boldsymbol{\epsilon}_{bj}. \quad (3.14)$$

The above formulae are useful when determining the relationship between the saturated (3.8 - 3.10) and the structural model (3.4 - 3.7). First, the equations

are solved in terms of the unstandardised (denoted by \*\*) estimates, that is, the coefficients of  $\mathbf{x}$  on both levels, the intercepts on the between level and the covariances of the the residual variables on both levels:

$$(3.15) \quad \mathbf{\Pi}_w^{**} = \mathbf{\Lambda}_w(\mathbf{I} - \mathbf{B}_w)^{-1}\mathbf{\Gamma}_w$$

$$(3.16) \quad \mathbf{\Pi}_b^{**} = \mathbf{\Lambda}_b(\mathbf{I} - \mathbf{B}_b)^{-1}\mathbf{\Gamma}_b$$

$$(3.17) \quad \boldsymbol{\mu}_b^{**} = \boldsymbol{\nu}_b + \mathbf{\Lambda}_b(\mathbf{I} - \mathbf{B}_b)^{-1}\boldsymbol{\alpha}_b$$

$$(3.18) \quad \begin{aligned} \boldsymbol{\Sigma}_w^{**} &= \text{Cov}(\boldsymbol{\epsilon}_w) = \text{Cov}(\mathbf{\Lambda}_w(\mathbf{I} - \mathbf{B}_w)^{-1}\boldsymbol{\xi}_w + \boldsymbol{\epsilon}_w) \\ &= \mathbf{\Lambda}_w(\mathbf{I} - \mathbf{B}_w)^{-1}\boldsymbol{\Psi}_w \left[ (\mathbf{I} - \mathbf{B}_w)^{-1} \right]^T \mathbf{\Lambda}_w^T + \boldsymbol{\Theta}_w \end{aligned}$$

$$(3.19) \quad \begin{aligned} \boldsymbol{\Sigma}_b^{**} &= \text{Cov}(\boldsymbol{\epsilon}_b) = \text{Cov}(\mathbf{\Lambda}_b(\mathbf{I} - \mathbf{B}_b)^{-1}\boldsymbol{\xi}_b + \boldsymbol{\epsilon}_b) \\ &= \mathbf{\Lambda}_b(\mathbf{I} - \mathbf{B}_b)^{-1}\boldsymbol{\Psi}_b \left[ (\mathbf{I} - \mathbf{B}_b)^{-1} \right]^T \mathbf{\Lambda}_b^T + \boldsymbol{\Theta}_b. \end{aligned}$$

For the categorical variables in the unrestricted model, the between level mean  $\mu_{bp}$  was restricted to 0 and the variance of the within level residual  $\epsilon_{wpij}$  was restricted to 1. In order to compare the parameters of the unrestricted and restricted models, the parameters of the restricted model 3.15 - 3.19 need to be standardised.

For standardising the estimates, one needs to weight the entries of the parameter vectors and matrices which are based on categorical observed variables. Let  $\mathbf{\Delta}_w$  be a  $p$ -dimensional diagonal matrix, with  $1/\sqrt{\Sigma_{wpp}^{**}}$  on the diagonal if the  $p$ -th entry is categorical, and 1 if not. Similarly, let  $\boldsymbol{\delta}_b$  be a  $p$ -dimensional vector with  $\mu_{bp}^{**}$  as the  $p$ -th entry if the  $p$ -th variable is categorical, and 0 otherwise. To obtain the standardised estimates of the thresholds as well as the parameters in 3.15 – 3.19, the following definitions apply:

$$\begin{aligned} \mathbf{a}_k^* &= \mathbf{\Delta}_w(\boldsymbol{\tau}_k - \boldsymbol{\delta}_b) \\ \boldsymbol{\mu}_b^* &= \mathbf{\Delta}_w(\boldsymbol{\mu}_b^{**} - \boldsymbol{\delta}_b) \\ \mathbf{\Pi}_w^* &= \mathbf{\Delta}_w\mathbf{\Pi}_w^{**} \\ \mathbf{\Pi}_b^* &= \mathbf{\Delta}_w\mathbf{\Pi}_b^{**} \\ \boldsymbol{\Sigma}_w^* &= \text{Cov}(\mathbf{\Delta}_w\boldsymbol{\epsilon}_w) = \mathbf{\Delta}_w\boldsymbol{\Sigma}_w^{**}\mathbf{\Delta}_w \\ \boldsymbol{\Sigma}_b^* &= \text{Cov}(\mathbf{\Delta}_w\boldsymbol{\epsilon}_b) = \mathbf{\Delta}_w\boldsymbol{\Sigma}_b^{**}\mathbf{\Delta}_w. \end{aligned}$$

Thus, the estimates of the categorical variables are divided by the standard deviation of their within error, and they are centered by subtracting their group means. Finally, the standardised estimates are united in  $\mathbf{s}^*$  in the same order as the unrestricted estimates in  $\mathbf{s}$ .

The WLS fit function is similar to the case of standard one-level structural estimation 2.7:

$$(3.20) \quad F_{\text{WLS}} = (\mathbf{s} - \mathbf{s}^*)'\mathbf{W}(\mathbf{s} - \mathbf{s}^*).$$

The final estimates are obtained by minimising 3.20 in terms of the parameters of the structural model presented in (3.4)–(3.7).

In practice, it is more common to use only the diagonal of the weight matrix,  $\mathbf{G}_0^{-1}$ , instead of  $\mathbf{G}^{-1}$ , which is less strict on the requirements of number of clusters and sample size (Hox 2010). This method, diagonal weighted least squares (DWLS), is found to result in relatively small bias, even if the observed variables are continuous. Depending on the correction method of the fit statistic  $F_{\text{DWLS}}$ , DWLS can be referred to as WLSM and WLSMV. Both lead to same estimates and standard errors but different goodness-of-fit values. WLSM uses a mean correction and WLSMV a mean and a variance correction of the fit statistic. (Hox et al. 2010.)

A multilevel simulation study by Hox et al. (2010) showed that the differences between ML and DWLS estimates are negligible. A full ML estimation should yet be chosen unless there is a violation of its assumptions such as non-normal data. They concluded that the two most important factors that influence the fit of a model are the number of clusters and the estimation method. Under conditions similar to the dataset in this study (100 clusters, average size 10), WLSMV was found to have the smallest average chi-square bias and the best estimation accuracy of the five estimation methods tested. Therefore, in the analysis of this thesis, the estimation is conducted with WLSMV.



## 4 Model Building

Constructing and testing a (multilevel) structural equation model involves several stages. Snijders & Bosker (1999) consider model specification one of the most difficult parts in multilevel statistical inference. A model needs to fulfill both substantive and statistical expectations. One wants to describe the variance of the observed data adequately but without unnecessary complications to maintain substantive interest. In this chapter, I present three important issues related to model building.

*Identification* implies whether it is possible to estimate the model in a statistical sense. As for SEM, there are many identification rules whereas the identification of multilevel models is yet an unresolved question. Nonetheless, it is an important one. In section 4.1, I present the existing tools for assessing the identifiability of multilevel models.

Besides constructing the theory-based latent structural model, building and assessing a number of comparison models is recommended. In a *forward* approach, the researcher begins with a null model that contains no latent structure and adds new effects one by one, proceeding from the within level to the between level. The *backward* approach begins with specifying the two-level model of interest and constraining any non-significant paths or loadings one by one. (Snijders & Bosker 1999). The specification procedure described in section 4.2 is a combination of these two approaches.

After a model is estimated, its overall fit needs a thorough assessing. The means for model fit assessment which are available for MSEM with the DWLS estimator are presented in section 4.3. If the model fits adequately, the focus may shift to the parameter estimates: whether they support the hypotheses, whether they are reasonable and whether their standard errors are relatively small. An estimator divided by its standard error is standard normally distributed, thus, its significance and reliability are easy to assess (Bollen 1989).

Sometimes an estimated model proposes impossible values for the parameters, such as a negative variance or a correlation greater than 1. Encountering this type of incidence, the model specification needs careful re-examining. If no misspecification is found and the disturbed value is relatively close to an acceptable value, it may be restricted. For example, a variance estimate of -0.01 may be constrained to 0. An anomaly might be a result from small variation within the sample due to a small sample size, or for some other reason. (Bollen 1989).

## 4.1 Identification

A structural equation model contains both known and unknown parameters. Each parameter needs to be identified. Identification in SEM refers to the ability to express a parameter in terms of the known elements in the model. The parameters which are known to be identified include the elements of  $\Sigma$ , that is, the sample variances and covariances. Since multilevel identification methods barely exist, the denotations and findings in this section are limited to the guidelines of Bollen (1989).

The unknown parameters of  $\mathbf{B}$ ,  $\Gamma$ ,  $\Phi$  and  $\Psi$  are gathered in  $\theta$ . If an unknown parameter in  $\theta$  can be written as a function of the elements in  $\Sigma$ , the parameter is identified. If all unknown parameters are identified, the model is identified. Thus, the problem of identification derives from the initial hypothesis  $\Sigma = \Sigma(\theta)$ . Another definition of identification requires that the equation  $\Sigma(\theta_1) = \Sigma(\theta_2)$  holds if and only if  $\theta_1 = \theta_2$ .

More often than not, the model needs some constraints to achieve identification. By constraints one usually means setting two or more parameters equal or determining some parameters fixed. For example, it is quite common to restrict the covariance of two variables to 0 if their relation is not relevant. If no universal scale for a latent variable exists, as is often the case with abstract constructs in social sciences, one of its factor loadings  $\lambda_{ij}$  can be restricted to 1. This way the fixed indicator becomes equal to the latent variable and the other loadings will be compared on the same scale.

There are two types of identification, *global* and *local*. Global identification refers to the above-mentioned description. Local identification implies the empirical means of examining identification, and it does not exclude the possibility that the parameters in  $\theta$  are not unique. Thus, local identification is easier to prove than global, but it lacks in power. A number of rules to simplify the global identification process have been developed for SEM but none for MSEM. In MSEM, the researchers still rely heavily on local identification. The two identification methods presented next define the global identification of SEM but can be applied to multilevel models as well.

*Algebraic solution* means manually solving the equation  $\Sigma = \Sigma(\theta)$ . It is a comprehensive identification rule that functions for simple models with few coefficients and variables. If all the unknown parameters can be written as a function of the known ones, the model is identified. A major drawback of the algebraic solution is that when the model is complex, the method becomes very complicated and inefficient.

The *t-rule* is a necessary rule but not a sufficient rule of identification. It means that if the condition is not met, the model is definitely not identified, while fulfilling the rule does not yet guarantee identification. Let  $t$  denote the number of free and unconstrained parameters in  $\theta$  and by  $p + q$  the number of observed variables in  $\mathbf{x}$  and  $\mathbf{y}$ . This leads to  $(\frac{1}{2})(p + q)(p + q + 1)$  equations to be solved in  $\Sigma = \Sigma(\theta)$ . If the number of unknown parameters is greater than the number of equations it needs to be solved from, it is clear that the model

is not identified. Thus, the t-rule is given by

$$t \leq \left(\frac{1}{2}\right) (p + q) (p + q + 1).$$

In multilevel models, the t-rule needs some modification before application. The number of free parameters includes the thresholds that are estimated from univariate distributions. As a consequence, their count needs to be subtracted from  $t$ . A challenge in multilevel models is that the models on different levels might not be specified in the same way, and identification on each level should be ensured.

## 4.2 Comparison Models

Even though the main interest of a researcher often lies in the fit of the structural model, it is recommended to test the fits of selected comparison models first. Assessing alternative models may widen the understanding of the variation in the data and also alleviate tracking potential misspecifications. In single-level SEM with ML estimation, information criteria and a  $\chi^2$ -difference test provide aid in testing whether the model has improved after the changes in the specification (Dilalla 2000). Neither is available for the comparison of nested multilevel structural equation models, which further complicates model specification.

There are two types of general comparison models. They are both relevant in the model fit assessment described in the next section. A *baseline* or *restricted* model contains no paths between the variables, and the only free parameters that are tested are the thresholds for the categorical variables and the residual variances. A *saturated* or *unrestricted* model, on the contrary, estimates all possible connections between the variables and it is a perfect reproduction of the sample data. All models containing the same group of variables are nested within the unrestricted model. The models which researchers estimate are often specified somewhere between the baseline and the saturated model. Consequently, it is valid to evaluate the fits of the target model and the two comparison models.

Instead of starting the specification from either the null model or the saturated model and strictly following the principles of forward or backward approach, Hox (2002) uses in his example a combination approach. It involves four main steps. After each step, a new between model is specified and the overall fit is evaluated. The within model is the structural model to be tested, and it remains unchanged.

First, a null model is specified. The model ignores the between level altogether. If the model holds, there might not be any between level structure in the data and proceeding with the multilevel model building may be unnecessary. If the model is insufficient in any way, one continues to the second step. The step involves specifying a between model with nothing else estimated but

the univariate variances. Hox (2002) calls this an independence model. If the independence model holds, there is considerable variation on the between level, but no substantially interesting structural model. The model is useful in the sense that it produces unbiased estimates of the individual model parameters.

In the third step, the researcher specifies a saturated model on the between level. The saturated model produces the best possible fit given the within model. Extending from Hox's approach, Ruy & West (2009) propose fitting also a saturated model on the within level to locate where a potential lack of fit might occur. Fourth and finally, one tests the structural model initially constructed. It may have the same specifications as the within level model, but also some separate components. One may test a few parallel between models. If two or more models have an equal fit, it is recommended to choose the simplest model that fits well for further interpretation.

### 4.3 Model Assessment

There are several ways for assessing the overall fit of a constructed model. Most are based on the minimum value of the fit function  $F$ . Methods to assess the models are divided into *absolute*, *relative* and *residual-based* indices. The first two compare the model to another, while the third measures the differences between the sample and the model estimated values. All the indices presented in this section are applicable to multilevel WLS estimation with non-normal observed variables. If the ML estimator is used instead, under multivariate normal conditions, the related likelihood function provides a variety of additional options for assessing the model fit.

Many of the model assessment properties are based on the minimum of the estimation function  $F(\mathbf{S}, \hat{\boldsymbol{\Sigma}})$ , denoted by  $\hat{F}$ . With WLS fit function,

$$(n - 1)\hat{F} \sim_{appr} \chi_d^2, \quad d = p(p + 1)/2 + p - q$$

where  $p$  is the number of observed variables and  $q$  is the number of free parameters in  $\theta$  (Yu 2002). Even though the  $\chi^2$  test is the most common model evaluation method in structural equation models, it has many drawbacks. To name but a few, it is extremely sensitive to sample size and kurtosis (Bollen 1989).

Diagonal weighted least squares (DWLS) does not share all the qualities of WLS, thus, the asymptotic distribution of  $\hat{F}$  is not a chi-square but a weighted sum of chi-square distributions with one degree of freedom (Asparouhov & Muthén 2010). As a result, the fit indices which are based on the asymptotic properties of the chi-square distribution require modification in order to be used and interpreted similarly to the WLS case.

The fit measures presented next are mainly indices of model fit, rather than actual statistical tests. Their cutoff criterion and power to detect model misspecifications vary between different types of models (Yu 2002). The cutoff values introduced in this section are mainly suggestive as MSEM models have

been studied little in general (Boulton 2011), and even less with the relatively recent DWLS estimation technique.

**Absolute fit indices** Absolute fit indices assess the overall model fit in terms of residuals, or compare the target model and the saturated model. The two examined in this thesis, the  $\chi^2$  test of the model fit and the root mean-square error of approximation (RMSEA), are based on sample discrepancy, that is, the difference between the unrestricted model and the estimated model.

To construct an equivalent of the ML- or WLS-based chi-square test to the case of DWLS estimation,  $\hat{F}$  needs to be modified. First, let  $D$  denote the difference in the number of parameters between the two models. Asparouhov & Muthén (2010) have proposed a new second order correction statistic  $\hat{F}^*$  which is chi-square distributed with  $D$  degrees of freedom, defined as

$$\hat{F}^* = \hat{F} \sqrt{\frac{D}{Tr(M^2)}} + D - \sqrt{\frac{DTr(M)^2}{Tr(M^2)}}$$

where  $M$  is a certain matrix defined by Satorra & Bentler (2001).

The RMSEA ranges from 0 to infinity, and it indicates the badness of fit. The higher the value, the larger the discrepancy. The RMSEA was introduced by Browne & Cudeck (1992) as

$$(4.1) \quad RMSEA = \sqrt{Max \left[ \left( \frac{\hat{F}^*}{D} - \frac{1}{n} \right), 0 \right]} = \sqrt{Max \left[ \left( \frac{\chi^2}{Dn} - \frac{1}{n} \right), 0 \right]}.$$

When the outcome variables are categorical, a function of sample variances replaces the degrees of freedom ( $D$ ) in 4.1 (Muthén 1998-2004). Values of the RMSEA close to 0 indicate a better fit. Two generally accepted thresholds are 0.08 for sufficient and 0.05 for good models (Browne & Cudeck 1992; Hu & Bentler 1999). The RMSEA is sensitive to the number of estimated parameters.

**Relative fit indices** The relative or comparative fit indices compare the estimated model ( $H_0$ ) to the baseline or restricted model ( $H_B$ ) and are sometimes referred to as indicators of goodness-of-fit. They measure how much of the information which is lost by fitting the restricted model is recovered by estimating the target model. (Boulton 2011).

Introduced by Bentler (1990), the comparative fit index (CFI) is defined as

$$(4.2) \quad CFI = 1 - \frac{Max(\chi_{H_0}^2 - D_{H_0}, 0)}{Max(\chi_{H_0}^2 - D_{H_0}, \chi_{H_B}^2 - D_{H_B}, 0)}.$$

The restricted model is nested within the estimated model. Thus, it is further from the optimal fit and the fraction in 4.2 is always less than 1. Hence, the range of the CFI is from 0 to 1 with higher values indicating a better fit.

The Tucker-Lewis index (TLI), named after its developers Tucker & Lewis (1973), is calculated by the following formula:

$$TLI = \frac{\frac{\chi_{H_B}^2}{D_{H_B}} - \frac{\chi_{H_0}^2}{D_{H_0}}}{\frac{\chi_{H_B}^2}{D_{H_B}} - 1}$$

The TLI and the CFI have the same range, and a recommended cut-off value of 0.95. A value less than 0.90 is considered unacceptable in both cases (Muthén 1998-2004; Hu & Bentler 1999; Yu 2002).

**Residual-based indices** Instead of comparing model fit, the residual-based indices measure the average differences between the sample ( $s_{jk}$ ) and the estimated ( $\hat{\sigma}_{jk}$ ) variances and covariances of the  $p$  observed variables in the model. The total number of covariances and variances is  $p(p+1)/2$ , denoted by  $e$ . The main developers of the following two indices presented next have been Muthén and Muthén (1998-2010).

The standardised root mean square residual (SRMR) is computed for each level separately, and it is defined for categorical outcomes as

$$(4.3) \quad SRMR = \sqrt{\frac{\sum_j \sum_{k \leq j} (s_{jk} - \hat{\sigma}_{jk})^2}{e}}$$

Hu & Bentler (1999) suggest a cut-off value of 0.08 for the SRMR. This index has been found sensitive to models with misspecification on factor covariances (Yu 2002). A related measure, the weighted root mean square residual (WRMR), weights the numerator in 4.3 with the estimated variances of the sample statistics vector. In the case of categorical variables, the WRMR takes the following form

$$WRMR = \sqrt{\frac{2n\hat{F}^*}{e}}$$

In her dissertation, Yu (2002) proposed a threshold of 1.0 for small samples ( $n=100$ ) and 0.95 for larger samples ( $n \geq 250$ ). The performance of the WRMR was found to depend on the normality of the observed variables and the sample size.

Fit indices to evaluate multilevel SEM are understudied. It is not given that the same statistical tests and goodness-of-fit indices which apply to single-level SEM are transferable to multilevel models. One MSEM simulation study with an ML estimator was conducted by Boulton (2011). His thesis concludes that, in many cases, a low ICC value indicates that the use of MSEM is questionable. He found that as the ICC increases, most tests (the  $\chi^2$  test, the RMSEA, the CFI, the TLI) become more sensitive to track misspecification on the between level of the model. The overall sample size increase was found to have the same

effect, as well as the average cluster size. Thus, the results from larger samples with homogenous clusters are more reliable than those from smaller samples with heterogenous groups.

Ruy & West (2009) introduced level-specific fit indices for MSEM to assess if a lack of fit in the model is due to discrepancy on solely one level of the model. These measures are based on the  $\chi^2$ -statistics of ML estimation and, consequently, cannot be applied to the introduced WLSMV method. In their study, Ruy & West (2009) conclude that the group level lack of fit remains often undetected when evaluating model fit on both levels in parallel. In this thesis, in the absence of a comprehensive set of level-specific fit indices, the emphasis is on the evaluation of the comparison models described in section 4.3.

## 5 Defining Poverty

The concept of poverty is more complicated than insufficient monetary income. Poverty is a complex, multi-dimensional phenomenon: gendered, dynamic, institutionally embedded, and also location specific. Poverty is often defined as 'the lack of what is necessary for material well-being', particularly regarding food, but also housing, land and other assets. Employment and psychological well-being can also be seen as dimensions of poverty. (Narayan, Patel, Schafft, Rademacher & Koch-Schulte 1999).

The need for measures capturing different aspects of welfare has been addressed in the development field (for example, Bourguignon & Chakravarty 2003; Ravallion 1996; World Bank 2001). One extensive framework which describes the nature of poverty more thoroughly than the strictly income based approach is the asset-based definition of livelihoods (Liverpool-Tasie & Winter-Nelson 2011; Carter & May 1999; UNDP 2004). It categorises the assets that define people's livelihoods into *natural*, *social*, *financial*, *human* and *physical* capital. The approach is very holistic, and, in practice, the different types of capitals are very difficult to quantify into just one or two indicators. When examining livelihoods based on these criteria, it is necessary to restrict the analysis to fewer aspects of the approach.

Assets and their relations are always local. Thus, site-specific research is needed in order to understand the nature of the problem in a certain time and a place (Mtapuri 2011). A major share of the research related to poverty is located in rural Africa. In South-East Asia, the livelihood issues have been studied mainly on meso (provincial or district) or macro level (Dasgupta et al. 2005; World Bank 2006). One problem encountered when examining poverty at higher levels is that the variation within the units is often greater than the variance between them, as governmental units such as provinces often consist of very different types of villages and households (World Bank 2006). World Bank (2006) also states that most relations between the dimensions of livelihoods should be analysed at the household level whenever possible. The community or village level is recommended when looking at the impacts of certain services that are part of the infrastructure accessible to the households.

The human development index (HDI) has been one attempt to capture the different dimensions of poverty in a single index. The HDI is a combination of national level aggregated indices on health, education and living standards. Health and education represent the human capital. The concept of living standards was designed to measure physical capital even though its indicator, gross



national income, is financial. (UNDP 2010). The HDI approach is aimed for international comparisons, and it does not enable exploring poverty experienced by individuals nor measuring inequalities within one country.

At the 20th anniversary edition of the Human Development Report in 2010, the publishing authority UNDP released international comparisons based on a new index, the multidimensional poverty index (MPI). Its aim was to address the weaknesses of the HDI while retaining its advantages. The MPI is based on the same three dimensions as the HDI but it is measured at the household level, and, thus, allows comparisons in any higher regional level. The indicator of living conditions is not based on income but the assets a household possesses.

In this chapter, I present how two-level structural equation modeling, the statistical method described in the three earlier chapters, can encompass the same dimensions of poverty as the MPI while featuring further relevant aspects which the MPI disregards. A national household survey from one of the poorest countries in the world, Lao PDR, is used as an example. The survey is referred to as the Finland Futures Research Centre (FFRC) survey from this on.

Section 5.1 presents the method behind the MPI in more detail with some results specific to Laos. After this, in section 5.2, I compare the indicators that were used to calculate the official MPI in Laos to the indicators available in the FFRC household survey. The comprehensive structural model is formed in the final section 5.3.

## 5.1 Multidimensional Poverty Index

A recent attempt to capture the dimensions of the asset-based definition of poverty in an applicable way has been the MPI, developed by Alkire & Foster (2007). The MPI includes three equally-weighted dimensions: health, education and living conditions. The dimensions are measured by ten indicators which have been chosen based on participatory exercises among the poor and data availability. Most of the indicators are linked to the Millennium Development Goals (MDGs), and, thus, are supported by international consensus. (Alkire & Santos 2010a.)

The method behind the calculation of the MPI is flexible and it can be applied to other subjects and levels. The method involves choosing the dimensions and their indicators, assigning weights to the indicators and determining cutoffs which categorise the observations. Each indicator is usually dichotomous, and their weighted sum is calculated for each observation. If the sum exceeds the selected cutoff, the unit is considered poor, or other distinct kind. The results are aggregated into higher level to receive an estimate of the occurrence of the incidence and the average intensity of the phenomenon. In the MPI, each dimension and indicator within a dimension has an equal weight but the number of indicators in a dimension varies. (Alkire & Foster 2007).

The MPI results for 104 developing countries were released in 2010 by UNDP. In the report, the MPI is presented as an advanced poverty measure,

to replace the international comparisons based on the HDI. The dimensions of poverty are the same in both but the observation level is different. In the MPI, it is essential to look at individuals or households and the data within a country needs to be from the same source, while the HDI consists of a set of aggregated indicators on national level. The MPI is aimed to explore inequalities also within regions and, indeed, it has revealed a large variation in poverty within countries. (Alkire & Santos 2010a).

A household deprived in terms of one third of the weighted indicators is identified as poor (Alkire 2011). In Laos, 47 % of the households are considered poor with this method. According to the traditional income-based poverty line of \$1.25 per day per person, 44 % are poor. The data in the study is part of the third round of international multiple indicator cluster surveys (MICS) from 2006. The three dimensions contributed to poverty in Laos somewhat equally: living standards 39 %, education 33 % and health 28%. In the developing countries, it is common that living conditions contribute the most. (Alkire & Santos 2010b.)

The Alkire-Foster approach bases on simple mathematics. In the analysis of this thesis, the perspective is different. Instead of forcing a cut-off and determining some households poor and others non-poor, MSEM is used to explore the variation of the dimensions of poverty. The weights are not assigned beforehand but estimated by the modeling technique. The relation between the MPI and monetary indicators is explored, as it has been appointed one of the key areas for future research by UNDP (2010). Other factors related to the topic are also included in the MSEM model. These are all considered relevant in the literature but their inclusion is not feasible in the MPI framework.

## 5.2 Indicators of MPI

The multidimensional poverty index is based on ten indicators of three dimensions. Similarly, ten indicators for the same dimensions were chosen for the factor structure of the MSEM model to be constructed in this thesis. In this section, these indicators are described in detail by their dimensions: health, education and living conditions. Appendix C summarises the comparison of the indicators in the two datasets. Table 5.1 provides the distributions of each indicator in the FFRC survey. Unfortunately, comparison data from MICS3 were not available.

Even though, in many cases, the indicators are not equivalent, they represent the same phenomena. The MPI was not designed as an exhaustive method but a flexible one. In fact, although the MPI indicators were mainly chosen based on data availability, as many as 40 % of the countries lacked at least one indicator. (Alkire & Santos 2010b.)

All the surveys from which the MPI was calculated are multistage stratified designs (Alkire & Santos 2010b). The data sample applied in this thesis was designed by the multistage probability proportional to size clustering method,

**Table 5.1.** Poverty indicators in the FFRC household survey (n=1564)

Dimension	FFRC Survey (2011)	Distribution (%)	
Health	Frequency of having fish, meat or eggs	Rarely	3.4
		1-2 days a week	21.1
		Several days a week	46.6
		Daily	28.9
		Three or more	0.3
		Two	1.0
Education	Highest school level household head has graduated from	One	7.4
		None	91.4
		None	26.2
		Primary School	46.0
		Secondary school	24.4
		Higher	3.5
Living Conditions	How many members can read or write	No one	2.7
		At least two cannot	10.2
		One cannot	27.4
		All	59.7
		No electrical lighting	16.6
Living Conditions	Access to latrine sanitation become worse in the past five years	House not completely made of permanent materials	25.8
		Wood as the main cooking fuel	80.2
		No car and no more than one of the following: bicycle, motorcycle, radio, fridge, telephone, television	24.9

stratified by region and village type. The capital of the country was excluded from the sample, as the main interest of the study was on rural poverty.

**Health** The most difficult dimension to measure proved to be health (Alkire & Santos 2010b). The MPI uses two indicators: undernourishment and child mortality. Definitions of undernourishment vary between surveys and countries. Child mortality has been particularly problematic, and it is missing altogether from the Laotian data 2006. All of the few missing responses in the dataset are related to undernourishment.

On the FFRC data, the households were asked how often they eat protein-containing food such as fish, meat or eggs. The replies were assumed to correspond to undernourishment. The second health-related indicator was the number of persistent health problems the household members had suffered from. Overall, 91 % claimed no one in their household had had any persistent health

problems. Although the figure sounds underestimated, it is in line with the results from the latest social and economic indicator survey (LECS) in 2007-2008 by the Government of Laos (2009). In this survey, 10 % of the respondents reported having had temporary health problems in the past four weeks. Only 2.3 % had suffered from long term illness or a disability.

**Education** In the MPI, education is measured by two indices. One criterion is that no one in the household has completed five years of schooling, and the other requires that at least one school-age child is not enrolled in school. These criteria can be criticised for not measuring the quality of education nor the level of knowledge attained (Alkire & Santos 2010b).

In the FFRC survey, the first education measure was the highest school level the household head had graduated from. Household headship is a recognised role within a family, and it is often assigned to a male. Considering that the household head is likely to be the most educated member in a household and that the completion of primary school requires five years of education, the first education indicators are close to equivalent.

Second question related to education was if the household members can read and write. This presumably indicates the knowledge gained by education which neither of the MPI indicators managed to capture. The question was asked separately of the household head, his or her spouse and others aged 15 years or more.

**Living Conditions** On total, there are six MPI indicators of living conditions. The first three are directly based on the MDG's and the latter three have strong grounds in the development literature (Alkire & Santos 2010b).

The first living condition indicator is having access to clean drinking water. In the FFRC survey, access to a number of different sources of drinking water was investigated for wet and dry season separately, and the following were considered safe: shared or public well with pump, private well, bottled pure drinking water, tap water and gravity flow system. The household had to have access to at least one of these sources on both seasons to satisfy the criteria.

The second indicator, improved sanitation, was slightly problematic as the survey did not include a question on the actual condition of sanitation. Instead, the households were asked if the access to latrine sanitation had improved, remained the same or become worse in the past five years. The households were classified in two: overall worsened access and improved or not changed access.

Wood and charcoal were considered 'dirty' cooking fuels in the MPI. Their usage was classified as fulfilling one of the indicators of living conditions. In the FFRC data, 99.6 % of the households said to use either of these as the primary source of energy for cooking. Thus, the indicator would not have provided any differences between the households. Fuel wood is often burned as such and it is considered less efficient in terms of energy usage. Burning wood indoors is also

related to respiratory health problems. Since 80 % of the Laotian households use wood as their primary cooking fuel, this was selected as an indicator of using dirty cooking fuel.

Electricity was seen as one of the keys in terms of living conditions (Alkire & Santos 2010b). In the FFRC survey, electrification was defined on the village level. Consequently, electrification contained no within variation essential in multilevel modeling. Some households, although outside the reach of the national electricity grids, still had access to minor sources of electricity such as local small-scale grids and battery lights. As electrical lighting can be seen as one of the primary advantages of electrification (Alkire & Santos 2010b), the electricity indicator was replaced with the use of an electric light bulb in this study.

Quality of housing bases on the floor material in the MPI. The FFRC data were more extensive: wall and roof materials were also recorded. 16 % of the households lived in a house with at least one of these elements made of so called soft or temporary materials (for example, thatch, bamboo and leaves). This was chosen as a more comprehensive indicator than merely the floor material.

Finally, one indicator was formed on the basis of a number of assets, each of them having a wide surrounding literature (Alkire & Santos 2010b). The condition was satisfied if the household did not own a car and possessed at most one of the following: radio, television, telephone, bicycle, motorbike and refrigerator. It was possible to replicate the condition with the FFRC data where 25 % of the households did not own more than one of the listed assets.

### 5.3 Structural Model of Poverty in Laos

After defining the poverty indicators, the structural model to be tested can be formed. In this section, I validate the further parts included in the model. The aim of the study is to explore the relation between multidimensional poverty and monetary income on household and village levels. UNDP (2010) defines this relation as one of the most important future research focuses. The preliminary analysis by Alkire & Santos (2010b) suggests that the MPI captures a distinct aspect of poverty that slightly overlaps with income. However, in most countries the comparison has been conducted on an aggregated level because of lack of available household data from one source.

I begin with the treatment of the poverty indicators ( $\eta_{w1ij}$  and  $\eta_{b1j}$  loading on  $\lambda_{wp1}$  and  $\lambda_{bp1}$ , respectively). The indicators and their relations were not the main interest of this study. A second-order factor construct, with poverty in the second order and three dimensions measured by two or six indices in the first order proved computationally demanding and difficult to interpret. The overall aim was to focus on the second order construct, multidimensional poverty. Therefore, the second-order indicators were parceled into one variable measuring each dimension (see figure 5.1). The distributions of these parceled indicators are provided in appendix D.

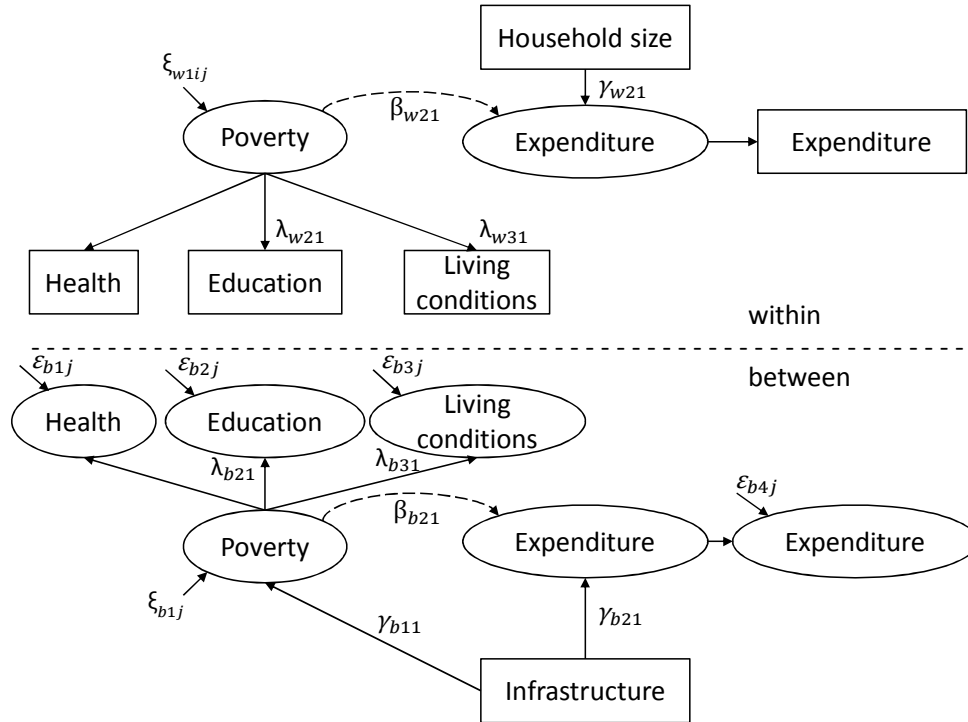
Besides multidimensional poverty, the second fundamental poverty indicator relates to financial issues. One can either measure the direct monetary income of a household, or its expenditure, which reflects consumption. Meyer & Sullivan (2003) argue that expenditure is more preferable than income because consumption measures material welfare more directly. They found that consumption indicates low material well-being better, especially for those with few resources. The Laotian households can be perceived as such. The World Bank (2001) considers expenditure estimates not only more reliable than current income in practice but also better in capturing long-run welfare levels. Hence, expenditure was chosen over income as the indicator of monetary resources.

In the FFRC survey, the expenditure was not asked as the net expenditure but in ordinal categories. In the model, it could have been specified as directly measured or as a single indicator latent factor ( $\eta_{w2ij}, \eta_{b2j}$ ) with 0 variance on the latent part ( $\psi_{w22}=0, \psi_{b22}=0$ ) and the loading fixed at 1 ( $\lambda_{w12}=1, \lambda_{b12}=1$ ). In estimation the two options are equivalent (section 2.2), and the latter was chosen for convenience.

The use of a single indicator restricts the available relationship types that can be modeled. The fewer indicators per latent factor, the more challenging it is to maintain identifiability. In the model studied in this thesis, the only attainable direct relation between multidimensional poverty and expenditure is a path from the former to the latter ( $\beta_{w21}, \beta_{b21}$ ).

Expenditure was measured as the total of the household. The net expenditure per household member is not calculable from the ordinary scale. Therefore, household size was added on the individual level as a variable explaining expenditure ( $x_{w1ij}$ , the effect on expenditure being  $\gamma_{w21}$ ). One of the MPI's two developers, Alkire (2011), states that one of the main weaknesses of the MPI method is its ignorance of household size. In some aspects, smaller households have a greater probability of being deprived. Alkire (2011) concludes that the overall effect of household size is not clear. Although the measures of multidimensional poverty in the structural model described here should not have bias due to household size, the path between the two ( $\gamma_{w11}$ ) was tested in one of the comparison models.

Access to public services is widely agreed an important non-market good relevant to welfare (Ravallion 2011). Consequently, the number of services in a village was chosen as an explanatory variable ( $x_{b1j}$ ) on the between level for both expenditure ( $\gamma_{b21}$ ) and multidimensional poverty ( $\gamma_{b11}$ ). The services cover most aspects of the asset-based theory of livelihoods briefly described in the beginning of this chapter: primary school, health clinic, electricity grid, transportation services and pagoda, to name but a few. Even though the services are not categorised in this model but treated as equally important, the most fundamental ones need to precede the less important. For example, a two-season road needs to be built before a petrol station can be constructed. Thus, the number of services in the village is assumed to indicate the level of infrastructure reasonably well.



**Figure 5.1.** The structural and measurement model for assessing poverty in Lao PDR. The main focus is on the relationships represented by the dashed arrows.

Finally, a word on the two-level approach. The World Bank (2006) recommends assessing poverty on the household level whenever possible. However, when looking at the impacts of certain services which are part of the infrastructure accessible to the households, the community or village level is advised. These suggestions are consistent with the model specified in this section.

Alkire & Santos (2010b) also consider groups keys in analysing the causes of multidimensional poverty. Their method for calculating the MPI allows studying group differences. Two-level SEM has, nevertheless, a notable advantage. The relations on the village and household levels can be studied simultaneously while separating the cluster and the individual effects. With MSEM, it is possible to compare the dependencies determined by the community and the dependencies the households themselves can influence. The nature of the relationships between the two poverty definitions is examined in the next chapter.

## 6 Results

Multilevel structural equation modeling, alike regular single-level SEM, is an extremely theory-orientated method. Models can be specified in a number of ways, and the variables can be modified, parceled or discarded altogether. With the new WLSMV technique presented, MSEM has become independent of one of its traditional assumptions, the normality of the observed variables. The methods becoming more and more flexible in different ways, it is increasingly important to have a clear guideline to follow when conducting the analysis. Otherwise the number of options may become overwhelming.

The basis for the specification of the structural model was presented and justified in the previous chapter 5. Now I present the results of the model building with the related background examinations, as introduced in chapter 4. The statistical theory behind the method was described in chapters 2 and 3. Thus, this chapter, presenting the study results, is founded on the contents of the previous chapters.

The data used on the analysis is described in section 6.1. The sample design and the considerations related to its hierarchical structure are examined. Section 6.2 provides an overview of the model specification process. Compared to the procedure described in chapter 4, the specification involves two additional saturated models for finding an optimal fit on both levels separately. The final model selected for a more detailed inspection is described in section 6.3, and its properties are discussed in the final section 6.4.

To date, there are no rules or methods for ensuring the identification of multilevel models. Identification is an assumption that enables estimating and interpreting the model but it does not provide any results in itself. In the lack of means for proving identifiability, a trivial example is presented in appendix E. The parameters on the between level are proven identified by solving them algebraically. A between level model with three paths was chosen as an example because it contains more parameters than the within level model by default when categorical variables are modeled. If the higher level construct is identified, the lower level model with more restrictions is also identified.

The observed covariance and correlation matrices are presented in tables 6.1 and 6.2. Like identification, the observed covariance structure is not a result as such even though it is the foundation of the method. The sample covariance (or correlation) matrix is what the structural model aims to replicate. The fit indices are generally based on the differences between the sample and the model estimated covariances.



**Table 6.1.** Sample within covariance/correlation matrix (n=1564). On the within level, the variances are scaled to 1 by default when the observed variables are ordinal.

	Health	Education	Living Conditions	Expenditure
Health	1.00	0.12	0.35	-0.21
Education	0.12	1.00	0.21	-0.17
Living conditions	0.35	0.21	1.00	-0.25
Expenditure	-0.21	-0.17	-0.25	1.00

By examining the covariance matrices, one can see that the observed variables have stronger dependencies on the between level (table 6.2) compared to the within level (table 6.1). Unfortunately, the significances of the correlations are not available. More profound inference is, therefore, provided as the final structural model is examined.

**Table 6.2.** Sample between covariance/correlation matrix (n=123). Below the diagonal: covariances, on the diagonal: variances, above the diagonal: correlations.

	Health	Education	Living Conditions	Expenditure
Health	0.70	0.85	0.65	-0.55
Education	0.42	0.35	0.78	-0.75
Living conditions	0.67	0.57	1.53	-0.72
Expenditure	-0.39	-0.37	-0.75	0.71

## 6.1 Sample

The sample of the Laos household data was designed by a multi-stage clustering method, with further stratification by region and village type. A number of households proportional to the village population were interviewed in 123 villages. In each village, from 4 to 18 (on average, 13) randomly selected households participated. There was also a separate questionnaire designed for the village head to obtain information on the village characteristics. The relevant parts of the household and the village head surveys were combined for the analysis of this thesis. Due to mismatches on the village identification variable, the original sample size of 1602 reduced to 1564 households.

Table 6.3 summarises the multilevel characteristics in the sample, as described in section 3.1. The intra-class correlations of the observed and parceled variables are all fairly high, ranging from 0.26 to 0.61. As a consequence, the design effects of each variable are also rather large (4.1 - 8.1), which means that the estimates have an increased chance of false significant results. However, the

risk can be considered minimal on the basis of the effective sample sizes. They range between 193 and 385, and, thus, are many times smaller than the actual sample size, 1564. Considering the large sample size, an acceptable significance level was determined to 1 %.

**Table 6.3.** Characteristics of the multilevel sample (n=1564, n<sub>clus</sub>=12.715)

Variable	Intra-class correlation	Design effect	Effective sample size
Health	0.411	5.81	269
Education	0.261	4.06	385
Living cond.	0.605	8.09	193
Expenditure	0.417	5.89	266

As a conclusion, the clusters in the sample are homogenous, which causes a threat to the reliability of the results (section 3.1) but, on the other hand, increases the sensitivity of the model fit indices to track misspecification (section 4.3). Large sample size on the lower level is essential in diminishing the large design effects, as measured by effective sample size. Sample size on the higher level, 123, substantially exceeds the size of 50 which Maas & Hox (2005) found to produce accurate and unbiased estimates. These arguments favour interpreting the MSEM results as they are estimated without further manipulation.

## 6.2 Model Comparison

In this section, the fits of the final model and the comparison models are evaluated. The fit index values of each model are summarised in table 6.4. The model on the individual level consists of expenditure  $\eta_{w2ij}$  regressing on the multidimensional poverty factor structure  $\eta_{w1ij}$  and household size  $x_{w1ij}$ , as illustrated in figure 5.1. This is the default within model specification apart from three exceptions. First, in the second between saturated model (2), the path from household size to multidimensional poverty ( $\gamma_{w11}$ ) is estimated. Second and third, full covariance matrices on the within level are estimated instead of structural relations in the two within saturated models (5, 6).

Beginning the model building with the null model (1) gave strong grounds to pursue further. Disregarding the multilevel nature of the data resulted in a non-positive definite covariance matrix which could not produce parameter estimates. Apart from the lack of estimates and a highly significant  $\chi^2$  value, the remaining fit index values - the RMSEA, the CFI, the TLI and the WRMR - were excellent. The next specified model was the independence model (2) with nothing but the univariate variances on the between level. The model was insufficient in terms of most of the fit indices (the RMSEA was tolerable, 0.073, and the within SRMR was good, 0.021).

**Table 6.4.** The fit indices of the comparison and parallel structural models

Model	value	$\chi^2$ df	p	Fit index					
				RMSEA	CFI	TLI	SRMR within	SRMR between	WRMR
1 Null	32.7	5	0.000	0.059	0.99	0.97	-	-	0.95
2 Independence	141.23	15	0.000	0.073	0.69	0.59	0.021	0.523	2.13
3 Saturated <sub>b</sub>	11.8	5	0.038	0.029	0.98	0.93	0.021	0.000	0.47
4 Saturated <sub>b</sub> + $\gamma_{w11}$	13.4	4	0.009	0.039	0.98	0.89	0.021	0.000	0.46
5 Saturated <sub>w</sub>	11.2	4	0.025	0.034	0.99	0.92	0.000	0.062	0.25
6 Saturated <sub>w</sub> - $\gamma_{b21}$	10.0	5	0.075	0.025	0.99	0.96	0.000	0.065	0.26
7 Final	21.4	10	0.018	0.027	0.98	0.95	0.020	0.062	0.55
Cut-off			0.01*	0.05 <sup>†</sup>	0.95*	0.95*	0.08 <sup>†</sup>	0.08 <sup>†</sup>	0.95 <sup>†</sup>

\*Greater than or equal to

<sup>†</sup> Smaller than or equal to

After excluding the possibility of a complete lack of multilevel structure, two saturated models on both levels were tested. One level being perfectly reproduced, misspecification on the other is easier to detect. Furthermore, level-specific optimal fits are easier to find. First, the saturated models were specified as suggested by the theory in section 5.3. Second, in the saturated models, one path was either added or removed. The model with the better one-level fit was chosen for the final model.

The first between level saturated model (3), constructed for testing the within level model, had an excellent fit in terms of every fit index examined. This indicated that there is no significant misspecification on the within level. The second between saturated model (4) tested the path from household size to multidimensional poverty ( $\gamma_{w11}$ ). The fit of the model worsened in terms of all of the indices apart from the SRMR and the WRMR. The  $\chi^2$  statistic and the TLI rejected the model. These values gave no reason to accept the tested path into the final model.

The first within saturated (5) model was specified on the between level as in figure 5.1. The multidimensional poverty factor structure  $\eta_{b1j}$  and expenditure  $\eta_{b2j}$  regressed on infrastructure  $x_{b1j}$ . The relationship between the latent poverty factors was causal, as expenditure regressed on multidimensional poverty. The model fit was good apart from the TLI being equal to 0.92 which was less than the recommended cut-off 0.95 but exceeded the ultimate threshold of 0.90. However, the path from infrastructure to expenditure was not significant ( $\gamma_{b21}=0.022$ , s.e.=0.033, p=0.51). This path was removed in the second within saturated model (6). The model fit improved substantially, and, consequently,  $\gamma_{b21}$  was restricted to 0 in the final model.

The final model (7) specified causal relationships on both levels. The overall fit is good, although the rather large value of the  $\chi^2$  statistic (21.4, df=10, p=0.018) is close to significant. It provided the only potential cause of a concern in the fit. The  $\chi^2$  test is known to reject models with large samples on the basis of arbitrarily small differences, and researchers often disregard it when analysing such data (Bollen 1989). The values of the RMSEA, the CFI, the TLI, both SRMR's and the WRMR are clearly in the favourable side of their respective cut-offs.

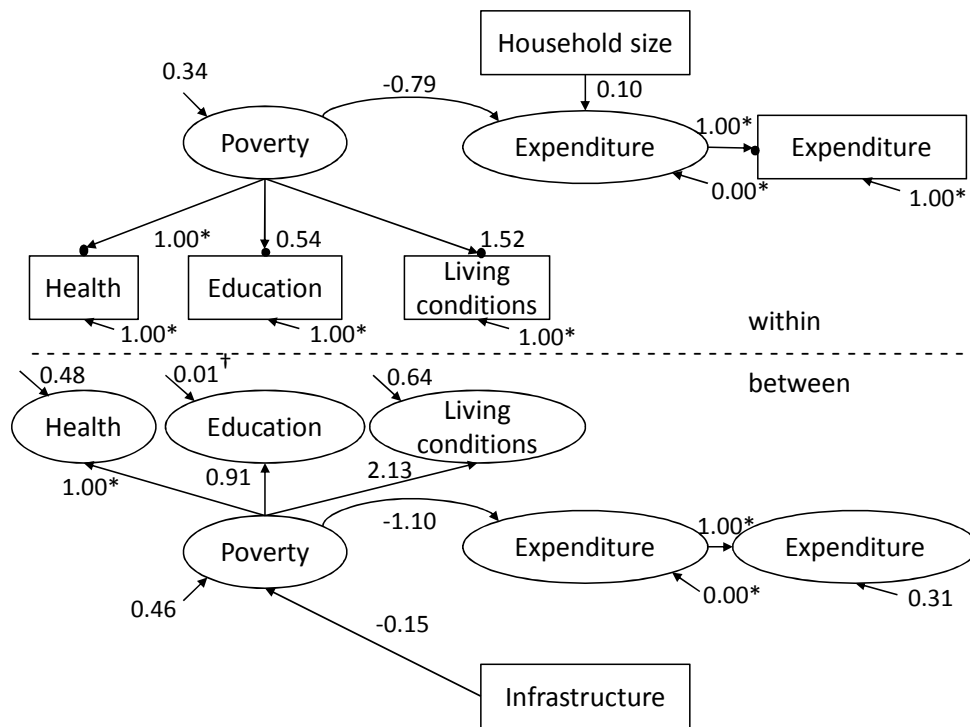
The parameter estimates in the final model are significant except for the between level residual variance of education. The parameter value is close to zero ( $\varepsilon_{b2j}=0.009$ ) with a relatively large standard error, resulting in a non-significant (p=0.81) estimate. Considering the good overall fit of the model, this issue was not subjected to further investigation nor was the model rejected.

### 6.3 Final Model

The final model (7), described in the previous section, was selected the best of the compared models based on existing theory and empirical testing. The parameter estimates are presented in figure 6.1. In the figure, the variables

expressed by squares are observed and the variables in ellipses are latent, standard normally distributed. A filled circle at the end of an arrow indicates that the variable being pointed to exists on both levels: on the within part, the variable consist of a measured part and a random intercept (cluster mean), and on the between level, the cluster means are latent variables which are modeled separately. One-way arrows express either factor indicators or regressed relationships. Arrows with an unenclosed beginning pointing to a variable indicate residual variances.

Fixed parameter values are expressed by a star (\*). All of the measured indicator variables were ordinal, and by default, their within level residual errors were fixed at 1 (subsection 3.2.1). Other given restrictions included assigning to the first indicator of each latent variable the loading of 1, and the constraints related to the single indicator latent variable expenditure (as explained in section 2.2).



**Figure 6.1.** The parameter estimates of the final model. All estimates but one are significant on 99 % confidence level.  $\chi^2=21.4$  (df=10, p=0.018), RMSEA=0.027, CFI=0.98, TLI=0.95, SRMR within=0.020 and between=0.062, WRMR=0.55. \*constrained †p $\geq$ 0.01

In large samples, the estimates divided by their standard errors are asymptotically normally distributed (Bollen 1989). Known as the Wald's test, the test is suitable for parameters which can take any value. Thus, its power is compromised when when assessing residual variances, which, naturally, are always positive.

Most estimates in the final model were strongly significant with p-values smaller than 0.001. The three exceptions were the residual variances of multidimensional poverty ( $\hat{\xi}_{b1j}=0.46$  with standard error 0.15 and their ratio 3.03,  $p=0.002$ ), living conditions (respectively,  $\hat{\epsilon}_{b3j}=0.64, 0.25, 2.61$  and 0.009) and, the already mentioned, education ( $\hat{\epsilon}_{b2j}=0.009, 0.036, 2.24$  and 0.81).

Health was assigned a scale variable in the factor structures and its loading was fixed at 1. Living conditions had the greatest and education the smallest loading of the three poverty indicators on both levels. On the village level, education loaded nearly twofold compared to the household level (0.91 and 0.54). Living conditions had also a larger loading on the village level, 2.13 compared to 1.52. The loadings indicate the expected change in the indicator if the value of the respective latent factor increases by 1.

As assumed, household size had a small but significant effect on household expenditure. If household size increases by one member, the expenditure is expected to increase by 0.10. On the between level, infrastructure had an equally significant but slightly stronger influence on multidimensional poverty. By adding one new element to the infrastructure of a village, multidimensional poverty is expected to decrease by 0.15. On the contrary to the initial hypothesis, the path from infrastructure to expenditure proved non-significant.

The regression coefficient from multidimensional poverty to expenditure is -0.79 on the within level. A negative relation was expected; decreases in the multiple dimensions of poverty increase monetary consumption. The respective coefficient on the higher level is almost 40 % greater, -1.10. Although infrastructure does not influence expenditure directly, it has a positive indirect effect on expenditure through multidimensional poverty, with the magnitude of  $-0.15 * (-1.10) = 0.17$ .

## 6.4 Discussion

Multilevel structural equation modeling provided valid and interpretable results about the nature of poverty in Laos. The final model, in its simplicity, unveiled features which cannot be examined by other statistical methods. Modeling dependencies on two levels simultaneously allows separating the effects that occur on one level alone.

Living conditions had the highest intra-class correlation and the strongest loadings on multidimensional poverty. If the level of poverty decreases, living conditions are most strongly affected. The result is consistent with the MPI study where living conditions contributed the most to the poverty measure. The ICC of education was the lowest, 0.26, which may have partly caused the estimated variance of almost 0 on the between level.

Household size proved not to have an impact on multidimensional poverty in Laos. Household size was added in the model to control for expenditure. The effect was fairly small but significant. Other assumptions on household characteristics that might impact poverty were not specified in the model. Although

the results suggest focusing on the community in reducing poverty, this study does not exclude the possibility that there are other household properties that advance poverty reduction equally or more effectively.

The services and infrastructure in the village proved an important factor behind both poverty measures. Multidimensional poverty regresses significantly on infrastructure and it mediates the effect to expenditure. The indirect effect of infrastructure on expenditure is, in fact, slightly stronger than its direct effect on multidimensional poverty. Some of these mediated relations are obvious. A hospital has a positive impact on health, which may or may not increase the average expenditure of the households through improved productivity. By having a local market in the village, the households have an opportunity to improve their living conditions by exchanging goods, which might, again, lead to increased expenditure.

Monetary measures assume that the markets for the goods and services relevant to one's well being exist, and they ignore the fact that if the service is provided for free, no monetary assets are needed (Ravallion 1996). This model was able to consider both of these arguments by including the three dimensions (infrastructure, expenditure and multidimensional poverty) in the same model. However, the types of infrastructure were not separated. Combining hospitals, schools, roads and others into one count variable does not reveal any specific impact a single service type might have. Some could be extremely effective while others insignificant.

The main interest of the study was the relationship between financial and multidimensional measures of poverty. The final model constructed showed that the relationship is significant on both household and village levels. However, on the village level it is considerably stronger. Poverty seems to arise from the surrounding village. Based on the final model, multidimensional poverty can be reduced most effectively by expanding the selection of services in the village.

Even though the multilevel design presumably did not cause biased or underestimated parameter values (section 6.1), parceling the indicator variables may have affected the error terms. Comparing the observed and the parceled distributions (table 5.1, appendix D) one can see that summing the items smoothed, at least to some extent, the distributions' skewness and kurtosis. This may have reduced the errors in the final model, which is typical when modeling parcels (section 2.2). As the loadings were not the main interest of the study, the risk was acknowledged and accepted.

Thus far, official information about the overlapping of income-poor and MPI poor is not available, nor other relevant research that would provide a contrast for the results of this study. Comparison to the results of the MPI is irrelevant. The focus in this study was not in classifying the households into poor and non-poor but to study the nature and relations of the poverty measures.

The statistical theory behind the model estimation was challenging to assimilate. Nevertheless, what was compromised in the efforts of studying the theory, was gained on the flexibility of applying the method. The model construction and specification, by using the program Mplus, proved simple and applicable,

and could be suggested to a researcher less familiar with statistical theories. A strong understanding about the model specification is, nevertheless, necessary. Contrary to many other programs developed for SEM, Mplus (version 6) does not offer a graphical representation of the model, such as, for example, figure 6.1. Another drawback is the fact that variables and their residual variances are not separated in the syntax but some commands are internally directed to the residual variances and others to the variable itself.

The initial intention was to explore more than one topic, poverty. Concepts such as 'energy poverty' (UNDP 2012) and poverty-environment nexus (see the World Bank 2006) were closer to the interest of the researcher and the project that funded the data collection. Lack of an extensive theory covering these topics and problems with the data led to discarding these ideas. Problems with the data may have been due to measurement errors: evaluating one's own energy usage per fuel and recent environmental changes is more difficult than evaluating the poverty indicators. The broader, mainly experimental models seldom converged. Thus, the focus shifted to poverty, which not only had a stronger theoretical background but also provided valid and interesting results.



## 7 Conclusions

The constructed two-level structural equation model explored different dimensions of poverty, how they are related, and how these relations differ on household and village levels. The final model was valid in terms of the indices assessing model fit. Some results were unexpected, while others confirmed existing theory.

The main interest was in the relation between monetary assets, measured by households' expenditure, and multidimensional poverty, whose indicators were adapted from the indicators of the multidimensional poverty index (UNDP 2010; Alkire & Santos 2010b). The nature of the relation had to be restricted to one-way paths from multidimensional poverty to expenditure to ensure statistical identification. The principal finding was that these two poverty measures have a significant negative causal relationship on both levels, although the relation is considerably stronger on the village level.

Household size had a weak positive impact on the household's expenditure but no significant influence on multidimensional poverty. Thus, poverty cannot be expected to diminish by simply reducing household size. A more effective factor was found on the village level. The number of different types of services and infrastructure had a significant direct effect on multidimensional poverty, whereas the effect on expenditure was indirect, mediated by multidimensional poverty.

As a conclusion, improving the infrastructure of the village influences poverty effectively. The result is important to take into consideration when planning and designing development aid projects. Lao PDR is part of the Mekong region, one of the target cooperation regions of Finnish development policy (Ministry of Foreign Affairs of Finland 2010). Recognising the influential role of the community, this study suggests that development aid projects in Laos should be targeted at villages with a poor level of services. Naturally, other means outside the scope of this study could prove equally or more effective.

The study was explorative in the sense that (multilevel) structural equation modeling has been barely used in social sciences. The results encourage MSEM to be pursued further on similar issues. This framework applied to Laos only. Conducting similar, possibly slightly modified analyses in other developing countries could provide an interesting platform for assessing location-specific characteristics of poverty. The data used in this analysis could also be further exploited. The sample was clustered in multiple stages enabling three- and four-level modeling, which could be tested once the software is complete.

# Bibliography

- Alkire, S. (2011), "Multidimensional Poverty and its Discontents", Working Paper 46, Oxford Poverty and Human Development Initiative, Oxford Department of International Development, University of Oxford.
- Alkire, S. & Foster, J. (2007), "Counting and Multidimensional Poverty Measurement", Working Paper 7, Oxford Poverty and Human Development Initiative, Oxford Department of International Development, University of Oxford.
- Alkire, S. & Santos, M. E. (2010a), "Multidimensional Poverty Index", Briefing 1, Oxford Poverty and Human Development Initiative, Oxford Department of International Development, University of Oxford.
- (2010b), "Acute Multidimensional Poverty: A New Index for Developing Countries", Working Paper 38, Oxford Poverty and Human Development Initiative, Oxford Department of International Development, University of Oxford.
- Asparouhov, T. & Muthén, B. (2007), "Computationally Efficient Estimation of Multilevel High-Dimensional Latent Variable Models", proceedings of the 2007 JSM meeting in Salt Lake City, Utah, *Section on Statistics in Epidemiology*, 2531–2535.
- (2008), "Growth Mixture Modeling: Analysis of non-Gaussian Random Effects", in *Longitudinal Data Analysis*, eds. Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G., Boca Raton: Chapman & Hall/CRC Press.
- (2010), "Simple Second Order Chi-Square Correction", Technical Appendix, Mplus Version 6.
- Bentler, P. M. (1990), "Comparative Fit Indices in Structural Models", *Psychological Bulletin*, 107, 238–246.
- Bollen, K. A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.
- Boulton, A. (2011), "Fit Index Sensitivity in Multilevel Structural Equation Modeling", MA thesis, University of Kansas.
- Bourguignon, F. & Chakravarty, S. R. (2003), "The measurement of multidimensional poverty", *Journal of Economic Inequality*, 1, 25–49.
- Browne, M. W. & Cudeck, R. (1992), "Alternative Ways of Assessing Model Fit", *Sociological Methods and Research*, 21, 230–258.
- Carter, M. R. & May, J. (1999), "Poverty, Livelihood and Class in Rural South Africa", *World Development*, 27, 1–20.
- Dasgupta, S., Deichmann, U., Meisner, C. & Wheeler, D. (2005), "Where is the Poverty-Environment Nexus? Evidence from Cambodia, Lao PDR and Vietnam", *World Development*, 33, 617–638.

- Dilalla, L. F (2000), "Structural Equation Modeling: Uses and Issues", in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, eds. Tinsley, H. E. & Brown, S. D., San Diego: Academic Press.
- Finland Futures Research Centre (2011) "INES - Interlinkages between Energy and Livelihoods, Laos Household Survey 2011", collector Indochina Research Ltd., producers Energy and Environment Partnership Mekong, Ministry for Foreign Affairs of Finland & Nordic Development Fund.
- Government of Laos (2009), *The Household of Lao PDR, Social and Economic Indicators: Survey results on Expenditure and Consumption of Household 2007/2008*, Vientiane: Ministry of Planning and Investment, Lao Department of Statistics.
- Hox, J. (2002), *Multilevel analysis: Techniques and Applications*, New Jersey: Lawrence Erlbaum Associates Inc.
- (2010), *Multilevel analysis: Techniques and Applications* (2nd ed.), New Jersey: Lawrence Erlbaum Associates Inc.
- Hox, J., Maas, C. & Brinkhuis, J. (2010), "The Effect of Estimation Method and Sample Size in Multilevel Structural Equation Modeling", *Statistica Neerlandica*, 65, 157–170.
- Hu, L. & Bentler, P. M. (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternative", *Structural Equation Modeling*, 6, 1–55.
- Jöreskog, K. G. (1990), "New Developments in LISREL: Analysis of Ordinal Variables Using Polychoric Correlations and Weighted Least Squares", *Quality and Quantity*, 24, 387–404.
- Lei, P-W. (2009), "Evaluating Estimation Methods for Ordinal Data in Structural Equation Modeling", *Quality and Quantity*, 43, 495–507.
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002), "To Parcel or not to Parcel: Exploring the Question, Weighing the Merits", *Structural Equation Modeling*, 9, 151–173.
- Liverpool-Tasie, L. S. & Winter-Nelson, A. (2011), "Asset Versus Consumption Poverty and Poverty Dynamics in Rural Ethiopia", *Agricultural Economics*, 42, 221–233.
- Maas, C. & Hox, J. (2005), "Sufficient Sample Sizes for Multilevel Modeling", *Methodology*, 1, 86–92.
- Maydeu-Olivares, A., García-Forero, C., Gallardo-Pujol, D. & Renom, J. (2009), "Testing Categorized Bivariate Normality with Two-Stage Polychoric Correlation Estimates", *Methodology*, 5, 131–136.
- McLachlan, G. J. & Krishnan, T. (1997), *The EM Algorithm and Extensions*, New York: Wiley.
- Meyer, B. D. & Sullivan, J. X. (2003), "Measuring the Well-Being of the Poor Using Income and Consumption", Working Paper 9760, National Bureau of Economic Research.
- Ministry of Foreign Affairs of Finland (2010), *Regional cooperation in the Mekong*, available at <http://www.formin.fi> (accessed 13.7.2012).
- Mtapuri, O. (2011), "Developing an Asset Threshold Using Consensual Approach:

- Results from Mashonoland West, Zimbabwe”, *Journal of International Development*, 23, 29–41.
- Muthén, B. O. (1984), ”A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators”, *Psychometrika*, 49, 115–132.
- (1997), ”Latent Variable Modeling of Longitudinal and Multilevel Data”, *Sociological Methodology*, 27, 453–480.
- (1998-2004), *Mplus Technical Appendices*, Los Angeles: Muthén & Muthén.
- Muthén, B. O. & Satorra, A. (1995), ”Complex Sample Data in Structural Equation Modeling”, *Sociological Methodology*, 25, 267-316.
- Muthén, L. K. & Muthén, B. O. (1998-2010), *Mplus User’s Guide: Statistical Analysis with Latent Variables* (6th ed.), Los Angeles: Muthén & Muthén.
- Narayan, D., Patel, R., Schafft, K., Rademacher, A. & Koch-Schulte, S. (1999), ”Can anyone hear us? Voices from 47 countries”, Volume 1, Voices of the Poor, Poverty Group, World Bank.
- Olsson, U. (1979), ”Maximum Likelihood Estimation of the Polychoric Correlation Coefficient”, *Psychometrika*, 44, 443–460.
- Ory, D. T. & Mokhtarian, P. L. (2010), ”The Impact of Non-Normality, Sample Size and Estimation Technique on Goodness-Of-Fit Measures in Structural Equation Modeling: Evidence from Ten Empirical Models of Travel Behaviour”, *Quality and Quantity*, 44, 427–445.
- Pearl, J. (2010), *Causality: Models, Reasoning and Inference* (2nd ed.), New York: Cambridge University Press.
- Raudenbush, S. W. & Bryk, A. S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, California: Sage Publications, Inc.
- Ravallion, M. (1996), ”Issues in measuring and modelling poverty”, *Economic Journal*, 106, 1328–1343.
- (2011), ”On Multidimensional Indices of Poverty”, Policy Research Working Paper 5580, The World Bank Development Research Group, the World Bank.
- Ruy, E. & West, S. G. (2009), ”Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling”, *Structural Equation Modeling*, 16, 583–601.
- Satorra, A. & Bentler, P. M. (2001), ”A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis”, *Psychometrika*, 66, 507–514.
- Şimşek, G. G. & Noyan, F. (2012), ”Structural Equation Modeling with Ordinal Variables: A Large Sample Case Study”, *Quality and Quantity*, 46, 1571–1581.
- Snijders, T. (2005), ”Power and Sample Size in Multilevel Modeling”, in *Encyclopedia of Statistics in Behavioral Science*, eds. Everitt, B. S., Howell, D. C., Chicester (etc.): Wiley.
- Snijders, T. & Bosker, R. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Analysis*, London: Sage Publications Ltd.
- Tucker, L. R. & Lewis, R. (1973), ”A Reliability Coefficient for Maximum Likelihood Factor Analysis”, *Psychometrika*, 38, 1–10.
- United Nations Development Programme (2004), *Attacking Poverty While Improving the Environment Nexus: Towards Win-Win Policy Options*, New York: United Nations Development Programme.

- (2010), *Human Development Report 2010: 20th Anniversary Edition*, New York: United Nations Development Programme.
- (2011), *Environment and Energy. Sustainable Energy*, New York: United Nations Development Programme.
- (2012), *Towards an 'Energy Plus' Approach for the Poor: A Review of Good Practices and Lessons Learned from Asia and the Pacific*, New York: United Nations Development Programme.
- World Bank (2001), *World Development Report 2000/2001: Attacking Poverty*, Washington, D.C: The World Bank.
- (2006), *Poverty Environment Nexus: Sustainable Approaches to Poverty Reduction in Cambodia, Lao PDR and Vietnam*, Washington, D.C: The World Bank.
- Yang, C., Nay, S. & Hoyle, R. H. (2010), "Three Approaches to Using Lengthy Ordinal Scales in Structural Equation Models: Parceling, Latent Scoring, and Shortening Scales", *Applied Psychological Measurement*, 34, 122–142.
- Yu, C-Y. (2002), "Evaluation Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes", Ph.D. thesis, University of California.

## Appendix A

# Expectation-Maximisation Algorithm

Let us denote the incomplete data where the unknown parameters  $\boldsymbol{\theta}$  ( $\in \Omega$ ) are estimated from as  $\mathbf{y}$ , and the complete but unobservable data as  $\mathbf{z}$ . Their sample spaces are  $S_y$  and  $S_z$ , and the relations from  $S_z$  to  $S_y$  are many-to-one. The probability function of  $\mathbf{y}$  is  $f_y(\mathbf{y}; \boldsymbol{\theta})$ , and similarly for  $\mathbf{z}$ . Instead of  $\mathbf{z}$ , one observes the incomplete vector  $\mathbf{y}=\mathbf{y}(\mathbf{z})$  that determines  $S_z(\mathbf{y})$ , a subset of  $S_z$ . Thus, the probability function of  $\mathbf{y}$  can be written as

$$(A.1) \quad f_y(\mathbf{y}; \boldsymbol{\theta}) = \int_{S_z(\mathbf{y})} f_z(\mathbf{z}; \boldsymbol{\theta}) d\mathbf{z}$$

The likelihood is estimated in terms of the likelihood function of the complete data, given the observed  $\mathbf{y}$ . That is,  $\log L_z(\boldsymbol{\theta})|\mathbf{y}$ . The algorithm starts by assigning some initial values  $\boldsymbol{\theta}^{(0)}$  for the parameters in  $\boldsymbol{\theta}$ .

The E-step calculates the expected likelihood in terms of  $\boldsymbol{\theta}^{(0)}$ , and its general form is

$$(A.2) \quad Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}} \{ \log L_z(\boldsymbol{\theta}) | \mathbf{y} \},$$

where  $k = 0, 1, \dots$  is the count of the iteration process.

In the M-step, one chooses any  $\boldsymbol{\theta}^{(k+1)}$  which maximises  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$  for all  $\boldsymbol{\theta} \in \Omega$ :

$$(A.3) \quad Q(\boldsymbol{\theta}^{(k+1)}; \boldsymbol{\theta}^{(k)}) \geq Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}).$$

The two steps are repeated until the difference in the likelihoods

$$(A.4) \quad L(\boldsymbol{\theta}^{(k+1)}) - L(\boldsymbol{\theta}^{(k)}) < c$$

with  $c$  expressing an arbitrarily small number defined beforehand.

(McLachlan & Krishnan 1997)

## Appendix B

# Maximum Likelihood EM Algorithm

If the variable follows the normal distribution, the first-order statistics can be estimated by a zero-dimensional integration (that is, integrating a constant function) as shown by Raudenbush & Bryk (2002). In the case of obtaining first-order statistics for categorical variables, the integration is one-dimensional.

In the EM algorithm, the likelihood function of the variable(s) is formed first. In the first step, some random estimates are given to the univariate and bivariate estimates. The likelihood is then maximised, and the estimates that maximise the The observed likelihood is, in this case, denoted as

$$(B.1) \quad L(\mathbf{y}_{wij}, \mathbf{y}_{bj}) = \prod_{j=1}^C \int \phi(\mathbf{y}_{bj}) \prod_{i=1}^{N_j} \left( \int \phi(\mathbf{y}_{wij}) d\mathbf{y}_{wij} \right) d\mathbf{y}_{bj}$$

where  $\phi$  is the univariate normal probability function. The integration is conducted numerically by adaptive and non-adaptive quadratures, which are based on approximating the continuous distribution with a categorical distribution. The integration interval is divided into  $R$  nodes, whose intervals are expressed as  $n_r$ . The density of each node is an approximation of the probability that the latent variable receives the values of the node, expressed in the following way

$$(B.2) \quad \int \phi(y) dy \approx \sum_{r=1}^R \frac{\phi(n_r)}{\sum_{i=1}^R \phi(n_i)}.$$

If adaptive integration is used, the nodes are concentrated in the area where the posterior distribution of the random effects,  $\mathbf{y}_{bj}$ , is non-zero. The likelihood is then approximated as

$$(B.3) \quad L(\mathbf{y}_{wij}, \mathbf{y}_{bj}) \approx \prod_{j=1}^C \sum_{r=1}^R Pr(y_{bj} = n_{brj}) \prod_{i=1}^{N_j} \left( \sum_{s=1}^S Pr(y_{wij} = n_{wsij}) \right)$$

where  $n_{brj}$  and  $n_{wsij}$  are the integration nodes of the between and within models. For the EM algorithm, the posterior distribution of  $\mathbf{y}_{bj}$  is computed as

$$(B.4) \quad p_{brj} = Pr(y_{bj} = n_{brj} | *)$$

$$= \frac{Pr(y_{bj} = n_{brj}) \prod_{i=1}^{N_j} \left( \sum_{s=1}^S Pr(y_{wij} = n_{wsij}) \right)}{\sum_{r=1}^R Pr(y_{bj} = n_{brj}) \prod_{i=1}^{N_j} \left( \sum_{s=1}^S Pr(y_{wij} = n_{wsij}) \right)}$$

and the conditional distribution of  $\mathbf{y}_{wij}$  as

$$(B.5) \quad p_{wsij|r} = Pr(y_{wij} = n_{wsij} | *, y_{bj} = n_{brj})$$

$$= \frac{Pr(y_{wij} = n_{wsij})}{\sum_{s=1}^S Pr(y_{wij} = n_{wsij})}$$

where the log-likelihood to be maximised for the complete data is as follows

$$(B.6) \quad \sum_{r=1, j=1}^{R, C} p_{brj} \log(Pr(y_{bj} = n_{brj})) + \sum_{r=1, s=1, j=1, i=1}^{R, S, C, N_j} p_{brj} p_{wsij|r} \log(Pr(y_{wij} = n_{wsij})).$$

Alternatively, methods such as accelerated EM algorithm (AEM) can be used in order to obtain faster convergence. Also other numerical integration methods are available for approximating B.2.



## Appendix C

### Comparison of the indicators in multidimensional poverty index and FFRC data

Dimension	MPI (2006)	FFRC Survey (2011)
Health	At least one malnourished member	Frequency of having fish, meat or eggs
	One or more children has died*	Number of persistent health problems
Education	No one completed 5 years of schooling	Highest school level household head has graduated from
	At least one school-age child not enrolled in school	How many members can read and write
Living Conditions	No electricity	No electrical lighting
	No access to safe drinking water	No access to safe drinking water
	No access to adequate sanitation	Access to latrine sanitation become worse in the past 5 years
	House has dirt floor	House not completely made of permanent materials
	Dirty cooking fuel (wood, charcoal)	Wood as the main cooking fuel
	No car and no more than one of the following: bicycle, motorcycle, radio, fridge, telephone, television	No car and no more than one of the following: bicycle, motorcycle, radio, fridge, telephone, television

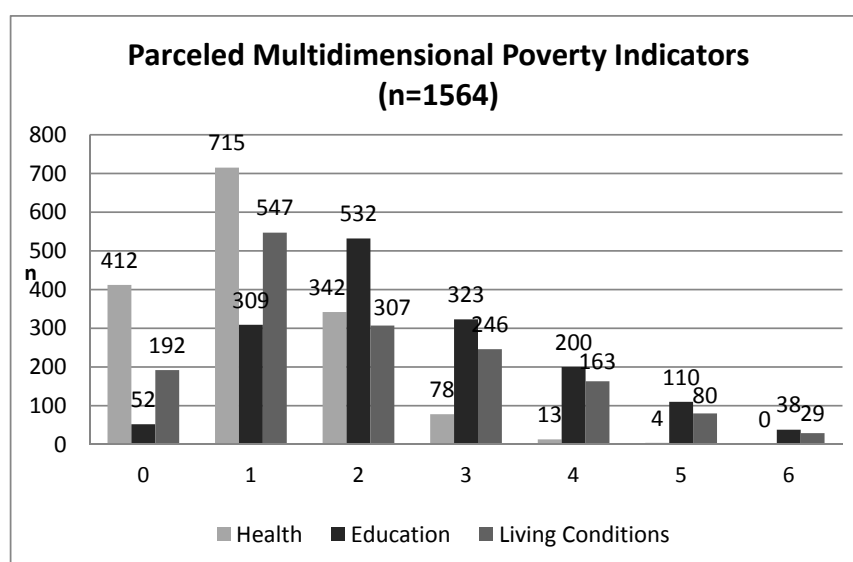
\*Missing from the MICS3 data in Laos, 2006

## Appendix D

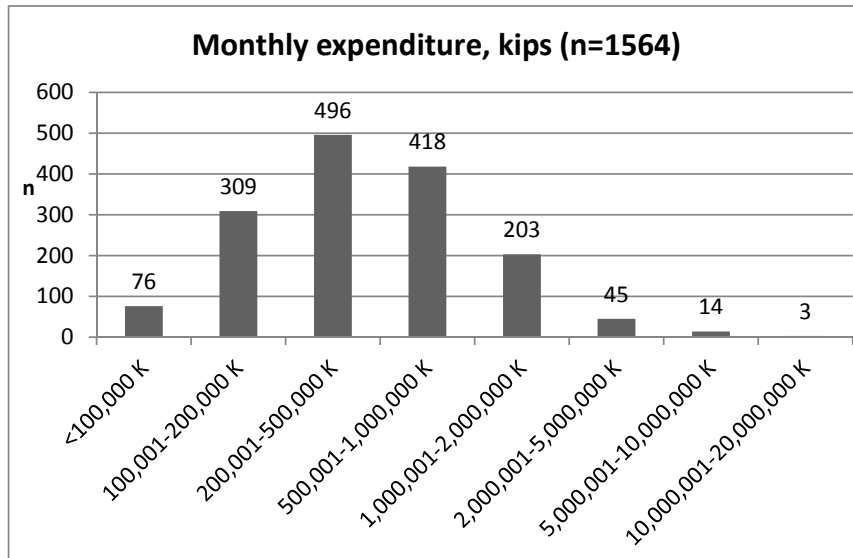
### Distributions of the measured variables in the MSEM model

**Table D.1.** Variables in the MSEM model

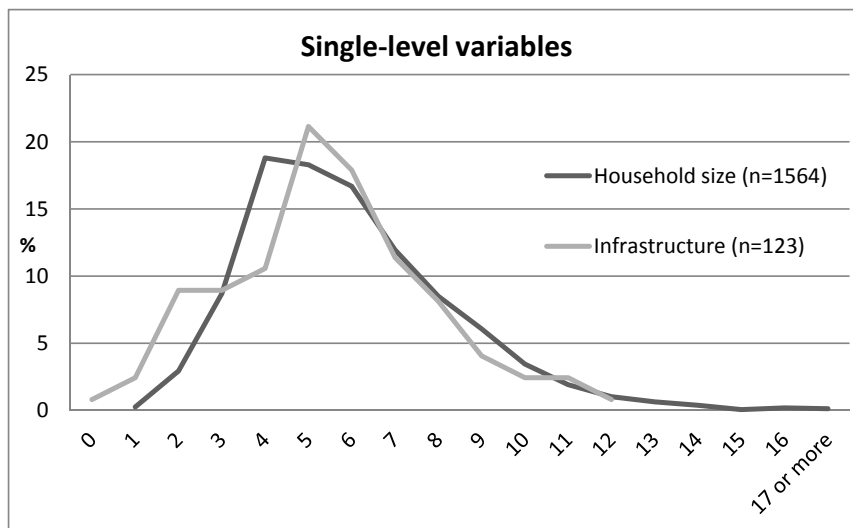
Variable	N	Mean	Median	Std. dev.	Variance	Min	Max
Health	1564	1.09	1	0.89	0.79	0	5
Education	1564	3.49	4	1.35	1.82	0	6
Living conditions	1564	2.00	2	1.47	2.17	0	6
Expenditure	1564	3.36	3	1.22	1.50	1	8
Household size	1564	5.93	6	2.39	5.70	1	22
Infrastructure	123	5.45	5	2.37	5.61	0	12



**Figure D.1.** The distributions of the multidimensional poverty indicators



**Figure D.2.** The distribution of monthly expenditure by category in kips



**Figure D.3.** The distributions of the variables that exist on a single level

## Appendix E

### Proof of the identifiability of a between level model using the algebraic solution

In this appendix, the identifiability of the between level model in comparison model 5 is proved by an algebraic solution. Following the denotations of subsection 3.2.1, the model is first written in matrix terms. Subscript  $i$  ( $i=1, 2, \dots, N_j$ ) denotes individuals,  $j$  ( $j=1, 2, \dots, 123$ ) clusters and  $p$  ( $p=1,2,3,4$ ) the observed dependent variables. The observed variable that exists only on the cluster level is  $x_{b1j}$ . The structural model is now as follows:

$$\begin{aligned}
 \mathbf{y}_{bj} &= (y'_{b1j} \ y'_{b2j} \ y'_{b3j} \ y'_{b4j}) \\
 &= \begin{pmatrix} \nu_{b1} \\ \nu_{b2} \\ \nu_{b3} \\ \nu_{b4} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \lambda_{b21} & 0 \\ \lambda_{b31} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{b1j} \\ \eta_{b2j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{b1j} \\ \varepsilon_{b2j} \\ \varepsilon_{b3j} \\ \varepsilon_{b4j} \end{pmatrix} \\
 &= \begin{pmatrix} \nu_{b1} \\ \nu_{b2} \\ \nu_{b3} \\ \nu_{b4} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \lambda_{b21} & 0 \\ \lambda_{b31} & 0 \\ 0 & 1 \end{pmatrix} \left\{ \begin{pmatrix} \alpha_{b1} \\ \alpha_{b2} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \beta_{b21} & 0 \end{pmatrix} \begin{pmatrix} \eta_{b1j} \\ \eta_{b2j} \end{pmatrix} + \begin{pmatrix} \gamma_{b11} \\ \gamma_{b21} \end{pmatrix} x_{b1j} + \begin{pmatrix} \xi_{b1j} \\ \xi_{b2j} \end{pmatrix} \right\} + \begin{pmatrix} \varepsilon_{b1j} \\ \varepsilon_{b2j} \\ \varepsilon_{b3j} \\ \varepsilon_{b4j} \end{pmatrix} \\
 &= \begin{pmatrix} \nu_{b1} \\ \nu_{b2} \\ \nu_{b3} \\ \nu_{b4} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \lambda_{b21} & 0 \\ \lambda_{b31} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{b1} + \gamma_{b11}x_{b1j} + \xi_{b1j} \\ \alpha_{b2} + \beta_{b21}\eta_{b1j} + \gamma_{b21}x_{b1j} + \xi_{b2j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{b1j} \\ \varepsilon_{b2j} \\ \varepsilon_{b3j} \\ \varepsilon_{b4j} \end{pmatrix} \\
 &= \begin{pmatrix} \nu_{b1} + \alpha_{b1} + \gamma_{b11}x_{b1j} + \xi_{b1j} + \varepsilon_{b1j} \\ \nu_{b2} + \lambda_{b21}(\alpha_{b1} + \gamma_{b11}x_{b1j} + \xi_{b1j}) + \varepsilon_{b2j} \\ \nu_{b3} + \lambda_{b31}(\alpha_{b1} + \gamma_{b11}x_{b1j} + \xi_{b1j}) + \varepsilon_{b3j} \\ \nu_{b4} + \alpha_{b2} + \beta_{b21}\eta_{b1j} + \gamma_{b21}x_{b1j} + \xi_{b2j} + \varepsilon_{b4j} \end{pmatrix}.
 \end{aligned}$$

The model contains 24 thresholds estimated from the univariate distributions and ten between level parameters to be solved:  $\lambda_{b21}$ ,  $\lambda_{b31}$ ,  $\gamma_{b11}$ ,  $\gamma_{b21}$ ,  $\beta_{b21}$ ,  $\text{var}(\xi_{b1j})=\psi_{b11}$ ,  $\text{var}(\varepsilon_{b1j})=\theta_{b11}$ ,  $\text{var}(\varepsilon_{b2j})=\theta_{b22}$ ,  $\text{var}(\varepsilon_{b3j})=\theta_{b33}$  and  $\text{var}(\varepsilon_{b4j})=\theta_{b44}$ . Other parameters in the model are intercepts or restricted (for example,  $\psi_{b22}=0$ ).

There are  $\frac{1}{2}(4+1+1)(4+1)=15$  known parameters in the sample covariance matrix and ten unknown parameters to be solved. Thus, the solutions provided

next are not unique. Let me begin with defining the sample variances and covariances. For simplicity, the uncorrelated items are already deleted from the formulas.

$$\begin{aligned}
\text{(E.1)} \quad \text{var}(y_{b1j}) &= \gamma_{b11}^2 \text{var}(x_{b1j}) + \psi_{b11} + \theta_{b11} \\
\text{(E.2)} \quad \text{var}(y_{b2j}) &= \lambda_{b21}^2 \gamma_{b11}^2 \text{var}(x_{b1j}) + \lambda_{b21}^2 \psi_{b11} + \theta_{b22} \\
\text{(E.3)} \quad \text{var}(y_{b3j}) &= \lambda_{b31}^2 \gamma_{b11}^2 \text{var}(x_{b1j}) + \lambda_{b31}^2 \psi_{b11} + \theta_{b33} \\
\text{(E.4)} \quad \text{var}(y_{b4j}) &= \text{var}(\beta_{b21} \eta_{b1j} + \gamma_{b21} x_{b1j} + \varepsilon_{b4j}) \\
&= \beta_{b21}^2 \text{var}(\eta_{b1j}) + \gamma_{b21}^2 \text{var}(x_{b1j}) + \theta_{b44} \\
&= \beta_{b21}^2 (\gamma_{b11}^2 \text{var}(x_{b1j}) + \psi_{b11}) + \gamma_{b21}^2 \text{var}(x_{b1j}) + \theta_{b44} \\
\text{var}(x_{b1j}) &= \text{var}(x_{b1j}) \\
\text{(E.5)} \quad \text{cov}(y_{b1j}, y_{b2j}) &= \text{cov}(\gamma_{b11} x_{b1j} + \xi_{b1j}, \lambda_{b21} (\gamma_{b11} x_{b1j} + \xi_{b1j})) \\
&= \lambda_{b21} \gamma_{b11}^2 \text{var}(x_{b1j}) + \lambda_{b21} \psi_{b11} \\
\text{cov}(y_{b1j}, y_{b3j}) &= \lambda_{b31} \gamma_{b11}^2 \text{var}(x_{b1j}) + \lambda_{b31} \psi_{b11} \\
\text{(E.6)} \quad \text{cov}(y_{b1j}, y_{b4j}) &= \text{cov}(\gamma_{b11} x_{b1j} + \xi_{b1j}, \beta_{b21} \eta_{b1j} + \gamma_{b21} x_{b1j}) \\
&= \text{cov}(\gamma_{b11} x_{b1j} + \xi_{b1j}, \beta_{b21} (\gamma_{b11} x_{b1j} + \xi_{b1j}) + \gamma_{b21} x_{b1j}) \\
&= \beta_{b21} \gamma_{b11}^2 \text{var}(x_{b1j}) + \beta_{b21} \psi_{b11} + \gamma_{b11} \gamma_{b21} \text{var}(x_{b1j}) \\
\text{cov}(y_{b2j}, y_{b3j}) &= \lambda_{b21} \lambda_{b31} \gamma_{b11}^2 \text{var}(x_{b1j}) + \lambda_{b21} \lambda_{b31} \psi_{b11} \\
\text{(E.7)} \quad \text{cov}(y_{b2j}, y_{b4j}) &= \text{cov}(\lambda_{b21} (\gamma_{b11} x_{b1j} + \xi_{b1j}), \beta_{b21} \eta_{b1j} + \gamma_{b21} x_{b1j}) \\
&= \lambda_{b21} (\beta_{b21} \gamma_{b11}^2 \text{var}(x_{b1j}) + \beta_{b21} \psi_{b11} + \gamma_{b11} \gamma_{b21} \text{var}(x_{b1j})) \\
\text{cov}(y_{b3j}, y_{b4j}) &= \lambda_{b31} (\beta_{b21} \gamma_{b11}^2 \text{var}(x_{b1j}) + \beta_{b21} \psi_{b11} + \gamma_{b11} \gamma_{b21} \text{var}(x_{b1j})) \\
\text{(E.8)} \quad \text{cov}(x_{b1j}, y_{b1j}) &= \text{cov}(x_{b1j}, \gamma_{b11} x_{b1j}) \\
&= \gamma_{b11} \text{var}(x_{b1j}) \\
\text{(E.9)} \quad \text{cov}(x_{b1j}, y_{b2j}) &= \lambda_{b21} \gamma_{b11} \text{var}(x_{b1j}) \\
\text{(E.10)} \quad \text{cov}(x_{b1j}, y_{b3j}) &= \lambda_{b31} \gamma_{b11} \text{var}(x_{b1j}) \\
\text{(E.11)} \quad \text{cov}(x_{b1j}, y_{b4j}) &= \text{cov}(x_{b1j}, \beta_{b21} \gamma_{b11} x_{b1j} + \gamma_{b21} x_{b1j}) \\
&= \beta_{b21} \gamma_{b11} \text{var}(x_{b1j}) + \gamma_{b21} \text{var}(x_{b1j})
\end{aligned}$$

The first coefficient of  $x_{b1j}$  and the factor loadings  $\lambda_{b21}$  and  $\lambda_{b31}$  can be solved directly from E.8 - E.10:

$$\text{(E.12)} \quad \gamma_{b11} = \frac{\text{cov}(x_{b1j}, y_{b1j})}{\text{var}(x_{b1j})}$$

$$\text{(E.13)} \quad \lambda_{b21} = \frac{\text{cov}(x_{b1j}, y_{b2j})}{\gamma_{b11} \text{var}(x_{b1j})} = \frac{\text{cov}(x_{b1j}, y_{b2j})}{\text{cov}(x_{b1j}, y_{b1j})}$$

$$\text{(E.14)} \quad \lambda_{b31} = \frac{\text{cov}(x_{b1j}, y_{b3j})}{\text{cov}(x_{b1j}, y_{b1j})}.$$

By utilising E.5 and E.12 - E.14, one obtains the variance of the latent residual

variable:

$$\begin{aligned}\psi_{b11} &= \frac{\text{cov}(y_{b1j}, y_{b2j})}{\lambda_{b21}} - \gamma_{b11}^2 \text{var}(x_{b1j}) \\ &= \frac{\text{cov}(y_{b1j}, y_{b2j})\text{cov}(x_{b1j}, y_{b1j})}{\text{cov}(x_{b1j}, y_{b2j})} - \frac{[\text{cov}(x_{b1j}, y_{b1j})]^2}{\text{var}(x_{b1j})}.\end{aligned}$$

The relation between  $\gamma_{b21}$  and  $\beta_{b21}$  can be solved from E.11:

$$(E.15) \quad \gamma_{b21} = \frac{\text{cov}(x_{b1j}, y_{b4j})}{\text{var}(x_{b1j})} - \beta_{b21}\gamma_{b11}.$$

By substituting  $\gamma_{b21}$  in E.6 by the form in E.15 and assuming  $\gamma_{b11}$  and  $\psi_{b11}$  known,  $\beta_{b21}$  can be identified.

$$\begin{aligned}\text{cov}(y_{b1j}, y_{b4j}) &= \beta_{b21}\gamma_{b11}^2 \text{var}(x_{b1j}) + \beta_{b21}\psi_{b11} \\ &\quad + \left( \frac{\text{cov}(x_{b1j}, y_{b4j})}{\text{var}(x_{b1j})} - \beta_{b21}\gamma_{b11} \right) \gamma_{b11} \text{var}(x_{b1j}) \\ \Leftrightarrow \text{cov}(y_{b1j}, y_{b4j}) - \text{cov}(x_{b1j}, y_{b4j})\gamma_{b11} &= \beta_{b21}(\gamma_{b11}^2 \text{var}(x_{b1j}) + \psi_{b11} - \gamma_{b11}^2 \text{var}(x_{b1j})) \\ \Leftrightarrow \beta_{b21} &= \frac{\text{cov}(y_{b1j}, y_{b4j}) - \text{cov}(x_{b1j}, y_{b4j})\gamma_{b11}}{\psi_{b11}}\end{aligned}$$

The second coefficient of  $x_{b1j}$  can be solved, for example, from E.7 where the other parameters have already been identified:

$$\begin{aligned}\gamma_{b21} &= \frac{\text{cov}(y_{b2j}, y_{b4j})}{\lambda_{b21}\gamma_{b11}\text{var}(x_{b1j})} - \frac{\beta_{b21}\gamma_{b11}^2 \text{var}(x_{b1j})}{\gamma_{b11}\text{var}(x_{b1j})} - \frac{\beta_{b21}\psi_{b11}}{\gamma_{b11}\text{var}(x_{b1j})} \\ &= \frac{\text{cov}(y_{b2j}, y_{b4j})}{\lambda_{b21}\gamma_{b11}\text{var}(x_{b1j})} - \beta_{b21}\gamma_{b11} - \frac{\beta_{b21}\psi_{b11}}{\gamma_{b11}\text{var}(x_{b1j})}\end{aligned}$$

Finally, the only parameters left unsolved are the residual variances of the  $y_{b_{pj}}$ 's,  $\theta_{b11}$ ,  $\theta_{b22}$ ,  $\theta_{b33}$  and  $\theta_{b44}$ , which are the only unknowns in E.1 - E.4. Consequently, they can be identified.