FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS
MASTER OF STATISTICS AND DATA SCIENCE
STRUCTURAL EQUATION MODELING

Assignment

Van den Broeck Sebastiaan (r0902562)

13/03/2023

1 Introduction

It has been noted that almost one in six individuals in the United States will experience a depressive disorder. Consequently, considerable personal, social and economic loss can be attributed to this type of illness. Although there is clearly a very large societal impact of depression, little is known about its relationship with personality and social functioning. In this work I will test a hypothesis related to depression that has been proposed by Tse et al. (2011). Specifically, they proposed that harm avoidance and self-directedness are indirectly linked to depression through social functioning. Moreover, there should also be a direct effect of self-directedness on depression. On the one hand, a behaviour can be classified under harm avoidance if it is done to avoid novelty and punishment. Self-directedness, on the other hand, is a form of self-determination and ability to regulate behaviour to suit goals and values. The authors have tested this hypothesis on a sample of university students, which limits the interpretability of their findings. By testing their hypothesis on a larger and more representative sample, I hope to contribute to the literature on depression. The dataset will be discussed next. Afterwards, a structural equation model has been used to test the hypothesis and will be discussed as well. Lastly, the results and implications thereof will be considered.

2 Preliminaries

2.1 Data

The data treated in this report is the Midlife in the United States (MIDUS) series. It is a national longitudinal study of health and well-being, created by a team of multi-disciplinary researchers. Currently, there are three waves in the study, which were collected via phone interviews, surveys and by bringing participants into clinical settings to facilitate collecting biological data. All three waves cover the contiguous United States in its entirety. The first wave was collected in 1995 and 1996, while the second wave was collected in 2004 and 2005. The most recent and third wave was collected in 2013 and 2014. In this analysis, the second and third wave have been combined to create a bigger dataset. It was not possible to incorporate the first wave, since a lot of variables changed between the first and second and third waves (Radler, 2014). In this section I will discuss the variables that have been used in the analysis.

An important reason for choosing this dataset is that it contains a lot of documentation for which variables form certain latent constructs such as depression or social anxiety. Since I am not too familiar with the field of psychology this would allow me to test a hypothesis that is better grounded in theory. First, depression is the most important latent variable in this work. It has been measured through seven questions during which the respondent reflects over the last two weeks. For example, the questions include losing interest, becoming tired, having trouble falling asleep or thinking about death. The responses have been recoded such that a higher score equates a higher level of depression. Specifically, each variable which measures this latent construct has been coded such that a 1 reflects a yes answer. As could be expected, a 0 then means a respondent has answered no.

Table 1: Depression indicators

Construct	Code	Question
Depression	PA63	During those two weeks, did you lose interest in most things?
	PA64	Thinking about these same two weeks, did you feel more tired out or low on energy?
	PA65	During those same two weeks, did you lose appetite?
	PA66	Did you have more trouble falling asleep than you usually do during those two weeks?
	PA67	During that same two week period, did you have a lot more trouble concentrating than usual?
	PA68	People sometimes feel down on themselves, no good, or worthless. During that two-week period, did you feel this way?
	PA69	Did you think a lot about death - either your own, someone else's or death in general - during those two weeks?

Table 2: Depression indicators distribution

Construct	Code	Count	
		0	1
Depression	PA63	126	479
	PA64	51	554
	PA65	263	342
	PA66	172	433
	PA67	88	517
	PA68	222	383
	PA69	229	376

Second, another important aspect in this report is harm avoidance. It has been described as an inheritable tendency for inhibiting behaviours to avoid novelty and punishment (Tse et al., 2011). Since it cannot be measured directly, four questions were asked to get an idea about this construct. First, interviewees were asked whether they would enjoy experiencing an earthquake or learning to walk the tightrope. These two variables were reverse recoded such that a 4 reflects not agreeing with the statement at all (harm avoidance), while a 1 indicates fully agreeing (no avoidance). Second, interviewees were presented with two scenario's twice. For each question, one scenario corresponds to a harmful situation, while the other scenario's is harmless. Again, there was a recoding such that a higher score on these two variables indicates avoiding harm.

Table 3: Harm avoidance indicators

Construct	Code	Question
Harm avoidance	SE7D	It might be fun and exciting to be in an earthquake.
	SE7V	It might be fun learning to walk a tightrope.
	SE8	Of these two situations, I would dislike more: Situation 1: Riding a long stretch of rapids in a canoe; Situation 2: Waiting for someone who's late.
	SE9	Of these two situations, I would dislike more: Situation 1: Being at the circus when two lions suddenly get loose down in the ring; Situation 2: Bringing my whole family to the circus and then not being able to get in because a clerk sold me tickets for the wrong night.

Table 4: Harm avoidance distribution

Construct	Code	Count			
		1 (harm)	2	3	4 (no harm)
Harm avoidance	SE7D	33	92	91	389
	SE7V	25	79	69	432

Table 5: Harm avoidance distribution

Construct	Code	Count	
		0 (harm)	1 (no harm)
Harm avoidance	SE8	335	270
	SE7V	276	329

Third, we should not forget about self-directedness, which has been measured through three variables. It evaluates the amount of self-determination and ability a respondent has in order to regulate behaviour to achieve goals and values (Tse et al., 2011). Making plans for the future, knowing what to want out of life and setting goals are important for this dimension. Again, the variables were reverse coded such that a higher score reflects agreeing more with the statement. The data indicates that most participants agree somewhat or fully what the three statements.

Table 6: Self-directedness indicators

Construct	Code	Question
Self-directedness	SE14O	I like to make plans for the future.
	SE14R	I know what I want out of life.
	SE14P	I find it helpful to set goals for the near future.

Table 7: Self-directedness distribution

Construct	Code	Count			
		1	**2**	**3**	**4**
Self-directedness	SE14O	48	140	241	176
	SE14R	67	144	235	159
	SE14P	47	149	228	181

Fourth, the latent variable social functioning has been used in the analysis. Seven questions related to this dimension were asked. The variables SE1BB, SE1D, SE1I and SE1V were reverse coded such that a higher score indicates a higher degree of social functioning. The dataset counts 601 observations after deleting rows with missing values. A missingness at random principle is therefore assumed.

Table 8: Social functioning indicators

Construct	Code	Question
Social functioning	SE1BB	People would describe me as a giving person, willing to share my time with others.
	SE1D	Most people see me as loving and affectionate.
	SE1HH	I have not experienced many warm and trusting relationships with others.
	SE1J	Maintaining close relationships has been difficult and frustrating for me.
	SE1P	I often feel lonely because I have few close friends with whom to share my concerns.
	SE1V	I enjoy personal and mutual conversations with family members and friends.

Table 9: Social functioning distribution

Construct	Code	Count						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Social functioning	SE1BB	4	7	16	36	70	207	265
	SE1D	5	14	23	68	74	225	196
	SE1HH	75	71	69	34	51	113	192
	SE1J	67	88	106	57	52	94	141
	SE1I	7	14	14	57	124	175	215
	SE1P	71	82	80	51	50	103	168
	SE1V	14	14	17	26	82	159	293

Lastly, the relationships between the variables themselves will be shortly considered. The correlation matrix is shown in Figure 1. The polychoric correlations play a pivotal role in this work and will be discussed in the next section. They have been shown in Figure 2 and are a little bit stronger than the traditional correlations. First, we may notice some triangles near the diagonal. This is a good sign that the indicators of a construct are correlated with each other. Convergent validity is important in the context of structural equation modeling and means that indicators which load on the same factor should be strongly related to each other (Brown, 2015). In the case of the present study, the indicators of the depression latent variable do not appear to be strongly correlated with each other. This may present a problem for the model, which will be further discussed later. Second, it is clear that the variables of self-directedness and social functioning are negatively correlated. This is a good sign, since the structural equation model will be able to capitalize on this relationship. Moreover, it appears that some of the depression and social functioning indicators are negatively correlated as well.

2.2 Invariance

Before continuing any further, it is important to establish the measurement properties of the model. Test bias, which occurs when items are not measuring the underlying constructs in the same way across groups, should be avoided at all costs. By simultaneously putting restrictions on multiple parameters of the measurement model, such equivalence can be tested (Brown, 2015). From the previous section it is clear that longitudinal data has been used. Are the changes in a construct then due to changes in the construct itself or due to changes in how the construct is measured over time? Moreover, other sources of test bias, specifically the sex and age of the participants, will be tested as well.

Configural invariance is the first step in this process and indicates that the factor structure is equal across groups. Second, metric invariance is tested by constraining the factor loadings to be equal

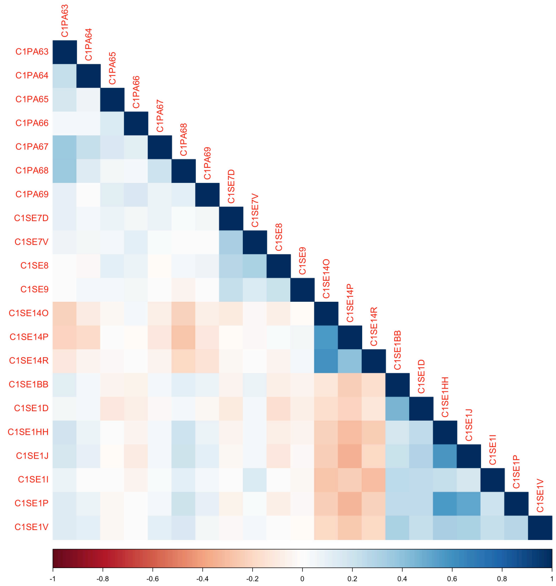

Figure 1: Correlation plot

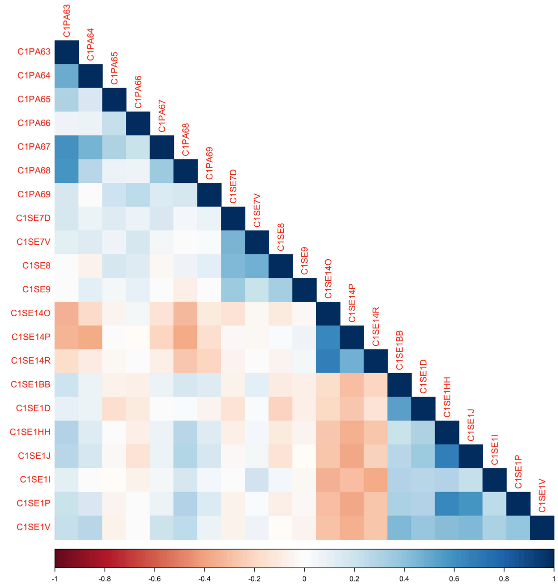

Figure 2: Polychoric correlation plot

across groups. Lastly, by also constraining the indicator intercepts to be equal, scalar invariance can be evaluated. A stepwise procedure can be employed by beginning with the least restricted solution and gradually testing whether the models χ^2 are significantly different from each other. Essentially, a likelihood ratio test is used: a rejection of the null hypothesis then indicates that invariance cannot be concluded and that the more restricted model has a worse fit (Brown, 2015). However, it should be noted that the χ^2 test is sensitive to sample size. Due to the large sample size in this study, other fit indices will therefore be used as well. The fit indices will be discussed in more detail later.

Table 10: Gender measurement invariance

Invariance form	χ^2	**df**	χ^2 **diff.**	**df diff.**	**p-value**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Configural	651.13	328	/	/	/	0.057	0.090	0.956	0.949
Metric	703.46	348	52.34	20	<0.001	0.058	0.092	0.952	0.947
Scalar	727.41	384	23.95	36	0.938	0.054	0.092	0.954	0.954

First, the invariance of the model with respect to gender will be tested. The χ^2 difference between the configural and metric model of 52.34 is itself χ^2 distributed with 20 degrees of freedom. In the metric model there are 20 parameters less to estimate, since there are 20 factor loadings which are constrained to be equal across the male and female group. Unfortunately, the p-value associated with the χ^2 test is highly significant ($p<0.001$), which indicates that the configural model has a better fit than the metric model. Hence, metric invariance cannot be concluded based on this test. Next, the χ^2 difference between the metric and scalar model of 23.95 is not significant ($p=0.938$). Based on this test, scalar invariance cannot be concluded, since metric invariance is still a necessary prerequisite. However, other fit measures should be taken into account due to the large sample size. The RMSEA, CFI and TLI (equations 5, 6, 7) directly make a correction for complexity through the degrees of freedom, while the SRMR indirectly makes a correction through the dimensionality of the reproduced covariance (correlation) matrix. It is therefore reasonable to expect the fit measures to be more or less the same for the configural, metric and scalar models if the invariance holds. Indeed, as shown in Table 10 this is the case. I will therefore conclude that scalar invariance for the male and female group holds.

Table 11: Time measurement invariance

Invariance form	χ^2	**df**	χ^2 **diff.**	**df diff.**	**p-value**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Configural	653.66	328	/	/	/	0.057	0.092	0.956	0.949
Metric	696.43	348	43.76	20	0.002	0.057	0.094	0.953	0.949
Scalar	717.76	384	21.33	36	0.975	0.054	0.093	0.955	0.956

Second, the two latest waves of the MIDUS dataset have been combined. The second wave was collected in 2004 and 2005, while the third wave originates from 2013 and 2014. This source of invariance should be investigated as well, because we want to be sure that changes in a construct are due to the construct itself and not because the measurement of the construct has changed over time. Again, we have to conclude that there is no evidence for metric and scalar invariance based on the χ^2 test. However, we now know that attention should be paid to other fit measures due to the large sample size. As shown in Table 11, the RMSEA, SRMR, CFI and TLI are very similar for the configural, metric and scalar models. I will therefore conclude that scalar invariance for the two waves hold, but it should be noted that the p-value associated with the χ^2 test is not significant.

3 The base model

Next, the base model will be discussed. Based on the work of Tse et al. (2011), it has been concluded that depression can be explained through the constructs harm avoidance, self directedness and social functioning. In other words, there are four latent variables which are related to each other through a structural model. Harm avoidance and self-directedness have an effect on social functioning. Social functioning, then, has an effect on depression. Moreover, it was estimated that there is also a direct effect of self-directedness on depression. Hence, there should be no direct effect of harm avoidance on depression.

In the previous section it became clear that ordinal variables have been used in the analysis. Model identification and estimation will therefore be discussed first. The measurement model indicates how variables are related to their latent constructs and will be discussed next. Afterwards, a closer look will be taken at the structural model, since an important aspect in this work is the relationships between latent variables. Lastly, the model fit will be evaluated through fit measures and further inspected using modification indices.

3.1 Model identification and estimation

First and foremost, model identification and estimation will be discussed. A structural equation model is said to be identified if every latent variable has its scale identified and the models degrees of freedom is zero or greater. To that end, the scale of the first indicator of every latent variable has been fixed to one. The model contains 85 parameters that should be estimated. Specifically, they are 16 (= 20 – 4) loadings, 4 regression parameters, 60 thresholds (more on this later), 1 latent covariance and 4 latent variances. The degrees of freedom of the model therefore equals 20*21/2 - 85 = 210 - 85 = 125. The following assumptions have been made on the equations shown in 1. The measurement errors δ are supposed to have an expected value of 0. It is assumed that they have constant variance across observations and are mutually uncorrelated. There should be a covariance of zero between these errors and the latent variables.

$$
\begin{cases}
\text{PA63} & = \lambda_{11}\text{depression} + \delta_{11} \\
\text{PA64} & = \lambda_{12}\text{depression} + \delta_{12} \\
\text{PA65} & = \lambda_{13}\text{depression} + \delta_{13} \\
\text{PA66} & = \lambda_{14}\text{depression} + \delta_{14} \\
\text{PA67} & = \lambda_{15}\text{depression} + \delta_{15} \\
\text{PA68} & = \lambda_{16}\text{depression} + \delta_{16} \\
\text{PA69} & = \lambda_{17}\text{depression} + \delta_{17} \\
\text{SE7V} & = \lambda_{21}\text{harm avoidance} + \delta_{21} \\
\text{SE7D} & = \lambda_{22}\text{harm avoidance} + \delta_{22} \\
\text{SE8} & = \lambda_{23}\text{harm avoidance} + \delta_{23} \\
\text{SE9} & = \lambda_{24}\text{harm avoidance} + \delta_{24} \\
\text{SE14O} & = \lambda_{31}\text{self-directedness} + \delta_{31} \\
\text{SE14P} & = \lambda_{32}\text{self-directedness} + \delta_{32} \\
\text{SE14R} & = \lambda_{33}\text{self-directedness} + \delta_{33} \\
\text{SE1BB} & = \lambda_{41}\text{social functioning} + \delta_{41} \\
\text{SE1D} & = \lambda_{42}\text{social functioning} + \delta_{42} \\
\text{SE1HH} & = \lambda_{43}\text{social functioning} + \delta_{43} \\
\text{SE1J} & = \lambda_{44}\text{social functioning} + \delta_{44} \\
\text{SE1P} & = \lambda_{45}\text{social functioning} + \delta_{45} \\
\text{SE1V} & = \lambda_{46}\text{social functioning} + \delta_{46} \\
\text{social functioning} & = \beta_1\text{harm avoidance} + \beta_2\text{self-directedness} + \delta_1 \\
\text{depression} & = \beta_3\text{social functioning} + \beta_4\text{self-directedness} + \delta_2
\end{cases}
\tag{1}
$$

A considerable problem arises when one considers the assumption of multivariate normality on the residuals δ. All observed variables are ordinal in nature, meaning that they are not continuous and should not be treated as such. Their means and (co)variances have no meaning, since they do not have origins or units of measurement (Jöreskog, 1994). It would therefore be questionable to take the usual approach in SEM of modeling the covariance matrix. The standard maximum likelihood machinery used in SEM is therefore also not applicable. In this case, robust maximum likelihood or a least squares approach (unweighted least squares, diagonally weighted least squares or weighted least squares) can be used (Yang-Wallentin et al., 2010). The method of diagonally weighted least squares has been specifically developed for ordinal data and has been shown to yield better results when the sample size is not small (Li, 2016). It has therefore been applied here. First, polychoric correlations are estimated. Afterwards, the model parameters can be estimated.

First, the polychoric correlations should be estimated. A solution can then be obtained by assuming that a latent, normal variable x^* is responsible for the observed ordinal variables x. With $x = m$ I mean to say that x belongs to a category m. Generally, the mean and variance if x^* are not identified, since only ordinal information is available (Şimşek and Noyan, 2012). They are therefore fixed to zero and one. Thresholds are used to link the latent variable to its observed counterpart:

$$x = m \text{ if } \nu_m < x^* < \nu_{m+1}. \tag{2}$$

Also, if one assumes x^* is standard normally distributed and ϕ and Φ denote the standard normal density and distribution functions:

$$\pi_m = Pr[x = m] = Pr[\nu_m < x^* < \nu_{m+1}] = \int_{\nu_m}^{\nu_{m+1}} \phi(u)du = \Phi(\nu_{m+1}) - \Phi(\nu_m). \tag{3}$$

In other words, a certain response m from the ordinal variable x is observed, if the response from its latent variable x^* falls between two thresholds. Hence, the thresholds are also parameters to be estimated. As far as I am aware, the polychoric correlations can then be estimated using maximum likelihood. Afterwards, the model parameters can be estimated using a weighted least squares approach. The ML and DWLS fit functions are defined as follows:

$$F_{ML} = \ln|S_{ML}| - \ln|\Sigma| + \text{trace}[(S_{ML})(\Sigma^{-1})] - p \qquad \text{(ML fit function)}$$

$$F_{DWLS} = [S_{DWLS} - \Sigma]' W_D^{-1} [S_{DWLS} - \Sigma]. \qquad \text{(DWLS fit function)}$$

S_{ML} is the covariance or correlation matrix and S_{DWLS} contains the polychoric correlations. Σ is the reproduced covariance or correlation matrix and depends on the models parameters. W_D^{-1} is a diagonal weight matrix, with weights that are inversely proportional to the variances of the polychoric correlations (Yang-Wallentin et al., 2010). We can therefore conclude that both approaches aim at minimizing the difference between a reproduced and sample covariance or correlation matrix. An important difference, however, is that the least squares approach allows a weighting for correcting for large polychoric correlations.

```
base.model <- "
    # measurement
    depression =~ C1PA63 + C1PA64 + C1PA65 + C1PA66 + C1PA67 + C1PA68 + C1PA69
    harm_avoidance =~ C1SE7V + C1SE7D + C1SE8 + C1SE9
    self_directedness =~ C1SE14O + C1SE14P + C1SE14R
    social_functioning =~ C1SE1BB + C1SE1D + C1SE1HH + C1SE1J + C1SE1P + C1SE1V

    # structural
    social_functioning ~ harm_avoidance+a*self_directedness
    depression ~ b*social_functioning + c*self_directedness

    IE := a*b
    TE := c + (a*b)
"

base.fit <- sem(base.model, data=data, ordered=TRUE, meanstructure=FALSE,
                estimator="DWLS")
summary(base.fit, standardized=TRUE, fit.measures=TRUE)
modindices(base.fit, sort=TRUE, maximum.number=20)
```

3.2 Measurement model

Second, we will take a closer look at the measurement model, which indicates how indicators relate to their latent constructs. The factor loading can be interpreted as the regression slope for predicting the indicator from the latent variable (Brown, 2015). The standardized loading is often more interesting, since it can be interpreted as a correlation and one does not need to worry about the scale of the variables. By squaring the standardized loading the communality can be obtained, which indicates the proportion of the variance in the indicator that is explained by the latent variable. The residual variance indicates the proportion of the variance that is not explained by the latent factor and therefore plays a pivotal role as well. Although there are no hard rules, a popular cut-off value for the communality appears to be 0.5 (Hair, 2010). Hence, more than half of the variance in the indicator should be explained by the latent variable. Based on my observation, communalities that are a little bit lower are also acceptable, as long as there is a good theoretical justification for the relationship between the factor and indicator. The standardized loading should then be larger than 0.7, which means that the indicator does a good job at reflecting the latent construct.

Inspecting Table 12, it is evident to see that the measurement model is lacking in some places. Using PA63, the respondent was asked about losing interest in most things. The high standardized loading of 0.850 indicates that the indicator is strongly correlated with its construct. The same applies to PA68 (participant feels down, no good or worthless), which has a communality of 0.572. In other words, 57.2% of the variance in this indicator is explained by the latent variable. Unfortunately, we have to conclude that the other depression indicators have a standardized loading that is too low. Consequently, they are weakly correlated with their construct and have a residual variance that is too high. The variables PA64, PA65, PA66, PA67 and PA69 are the problematic cases. In fact, the loading associated with PA66 is not even significantly different from zero ($p = 0.136$). The indicators PA64, PA65 and PA66 assess feeling low on energy, a loss of appetite and trouble falling asleep. PA67 and PA69 evaluate trouble concentrating and often thinking about death. A simple way to improve the model fit may be to reduce the number of variables that load on depression by deleting these problematic indicators. However, this action would lead to a decline of the theoretical support and validity of the model as well (Hair, 2010).

Next, the harm avoidance construct, which measures whether a behaviour is done to avoid novelty and punishment, plays a central role in this work. It is measured using four indicators: SE7V, SE7D, SE8 and SE9. In SE7V the participant is asked whether or not it might be fun to experience an earthquake. By squaring the standardized loading of 0.602, a low communality of 0.362 is obtained. In other words, there is a moderate correlation between the indicator and the harm avoidance construct, but the residual variance is too high. The variable SE7D was formed by asking the interviewee whether or not it might be fun to learn to walk the tightrope. The standardized loading of 0.724 is high enough. Next,

Table 12: Measurement model

Variable	Loading	Standard error	z-value	p-value	St. loading	Communality	Unique var.
PA63	1.000				0.850 (λ_{11})	0.722	0.278 (δ_{11})
PA64	0.631	0.073	8.662	<0.001	0.536 (λ_{12})	0.287	0.713 (δ_{12})
PA65	0.198	0.047	4.236	<0.001	0.168 (λ_{13})	0.028	0.972 (δ_{13})
PA66	0.071	0.048	1.492	0.136	0.061 (λ_{14})	0.004	0.996 (δ_{14})
PA67	0.646	0.066	9.757	<0.001	0.548 (λ_{15})	0.301	0.699 (δ_{15})
PA68	0.890	0.078	11.460	<0.001	0.757 (λ_{16})	0.572	0.428 (δ_{16})
PA69	0.360	0.051	7.005	<0.001	0.306 (λ_{17})	0.094	0.906 (δ_{17})
SE7V	1.000				0.602 (λ_{21})	0.362	0.637 (δ_{21})
SE7D	1.203	0.160	7.504	<0.001	0.724 (λ_{22})	0.525	0.475 (δ_{22})
SE8	1.194	0.155	7.718	<0.001	0.719 (λ_{23})	0.518	0.482 (δ_{23})
SE9	0.740	0.106	7.004	<0.001	0.446 (λ_{24})	0.199	0.801 (δ_{24})
SE14O	1.000				0.821 (λ_{31})	0.675	0.325 (δ_{31})
SE14P	0.982	0.049	19.988	<0.001	0.807 (λ_{32})	0.651	0.349 (δ_{32})
SE14R	0.851	0.043	19.704	<0.001	0.699 (λ_{33})	0.488	0.512 (δ_{33})
SE1BB	1.000				0.564 (λ_{41})	0.318	0.682 (δ_{41})
SE1D	0.956	0.056	17.145	<0.001	0.539 (λ_{42})	0.291	0.709 (δ_{42})
SE1HH	1.398	0.066	21.183	<0.001	0.788 (λ_{43})	0.621	0.379 (δ_{43})
SE1J	1.345	0.064	21.088	<0.001	0.759 (λ_{44})	0.575	0.425 (δ_{44})
SE1P	1.323	0.064	20.568	<0.001	0.746 (λ_{46})	0.556	0.444 (δ_{46})
SE1V	1.113	0.060	18.606	<0.001	0.628 (λ_{47})	0.394	0.606 (δ_{47})

SE8 and SE9 are used to measure harm avoidance by presenting the respondent with two situations. Afterwards, a choice has to be made between them. The standardized loadings are, respectively, 0.729 and 0.446. In the first question the harmful situation is riding a long stretch of rapids in a canoe, while the safe situation is waiting for someone who is late. In the second question the harmful situation is being at the circus when two lions get loose. The safe situation is bringing the family to the circus, but not being able to get in. The last question therefore not really presents a choice between a harmful and a safe situation, but rather a choice between a harmful and an embarrassing situation. Perhaps this could be the reason why the standardized loading is lower than in SE9.

Third, self-directedness has been described as a form of self-determination and ability to regulate behaviour to suit goals and values. Moreover, it should be related to the harm avoidance construct (Tse et al., 2011). The variables SE14O, SE14R and SE14P are used to measure this latent variable. SE14O evaluates whether or not the respondent likes to make plans for the future. The standardized loading of 0.821 indicates that there is a high correlation between the indicator and its latent variable. SE14R is used to measure whether the participant knows what he or she wants out of life. Again, the standardized loading of 0.807 is high enough. Lastly, SE14P is used to evaluate whether the participant finds it helpful to set goals for the near future. The standardized loading of 0.699 is a bit lower, but still high enough nonetheless.

Lastly, social functioning plays an important role, since it is believed to have a direct effect on depression (Tse et al., 2011). Seven indicators have been used to measure this latent variable. Using SE1BB the respondent was presented with a self-evaluation statement: 'People would describe me as a giving person, willing to share my time with others.' The standardized loading of 0.564 indicates a poor correlation and a high residual variance. The same problem is present in SE1D, which has a standardized loading of 0.539. SE1HH is used to measure whether the respondent has experienced many warm and trusting relationships with others. The communality of 0.788 indicates that there is a high correlation and the indicator does a good job in explaining the variance of the latent variable. In SE1J it was asked whether the interviewee has experienced difficulties in maintaining close relationships. The standardized loading of 0.788 is high enough and indicates that a one standardized unit increase in social functioning is estimated to lead to an increase of 0.788 standardized units in SE1J. Next, SE1P is used to measure whether the respondent often feels lonely because he or she has few close friends with whom to share his or her concerns. The standardized loading of 0.746 is high enough. SE1V is used to evaluate whether the participant enjoys personal and mutual conversations with family members and friends. The loading is borderline (0.628), but personally I would still consider it high enough.

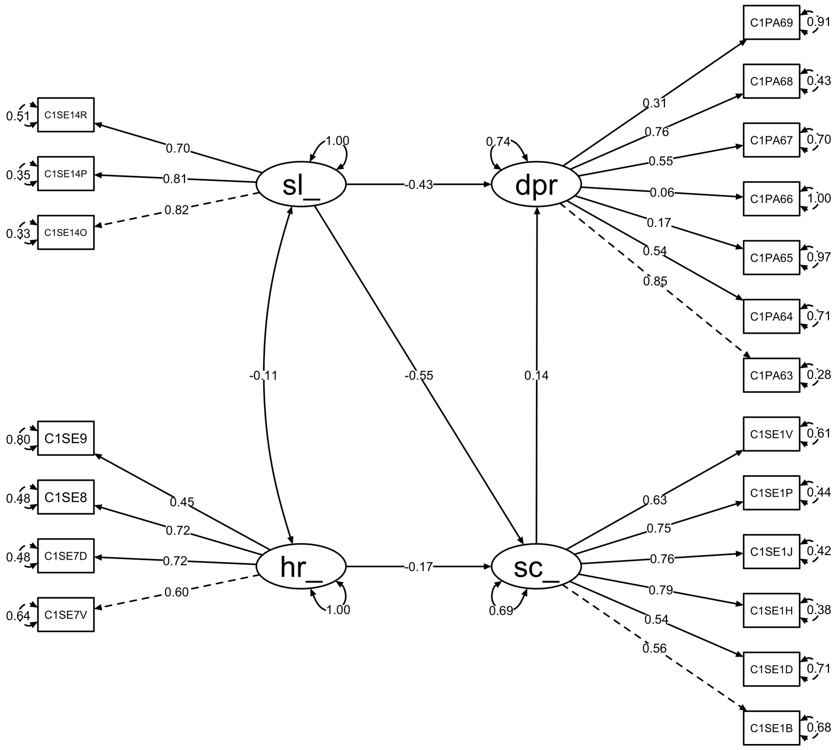

Figure 3: Summary of the (standardized) base model. hr_: harm avoidance, sl_: self-directedness, sc_: social functioning, dpr: depression

To sum up, there are a fair amount of indicators that have a low standardized loading. Some authors would therefore delete them from the dataset and refit the model. However, I have decided to leave them in the model, since there is a good theoretical justification for their relationship with the latent variable. Furthermore, one needs to take this problem into account when interpreting the results from the structural model and model evaluation. A chain is only as strong as its weakest link and in a structural equation model the measurement model is a very important link.

3.3 Structural model

The structural model is of great interest in this work, since it allows us to make conclusions about the relationships between the latent constructs. First, we may consider the direct effect of harm avoidance (β_1) and self-directedness (β_2) on social functioning. Both parameters estimates are highly significant and indicate a negative relationship with social functioning. On the one hand, it is estimated that there is a weak, negative relationship between harm avoidance and social functioning. The standardized regression coefficient of -0.172 indicates that an increase of one standardized unit in harm avoidance will lead to a decrease of 0.172 standardized units in social functioning. On the other hand, it appears that there is a stronger relationship between self-directedness and social functioning. This result is a bit counterintuitive, since the standardized parameter of -0.548 indicates a strong negative relationship. I would have personally expected a positive relationship between the two constructs.

Following the theory proposed by Tse et al. (2011), we may expect there to be a significant effect of social functioning on depression. Indeed, the results indicate that there is a negative pattern between the two in the sample, and this can be generalized to the population ($p < 0.001$). However, the standardized coefficient of 0.206 indicates a positive relationship, which is counterintuitive. In the literature, it is often stated that there is a negative relationship between the two (Tse et al., 2011). Lastly, we may consider the direct and indirect effect of self-directedness on depression. The standardized coefficient of -0.427 indicates a strong negative direct relationship between the two ($p < 0.001$). In other words, more self-directedness is estimated to lead to less depression. Next, the indirect effect of -0.075 is calculated by multiplying the standardized parameter estimates of the direct effect of self-directedness on social functioning ($\beta_2 = -0.548$) and the direct effect of social functioning on depression ($\beta_3 = 0.136$). The total effect (-0.502) can be calculated by adding the direct effect ($\beta_4 = -0.427$) to the indirect effect. The indirect, direct and total effects are all highly significant ($p < 0.001$).

$$\begin{cases} \text{social functioning} & = \beta_1 \text{harm avoidance} + \beta_2 \text{self-directedness} + \delta_1 \\ \text{depression} & = \beta_3 \text{social functioning} + \beta_4 \text{self-directedness} + \delta_3 \end{cases} \tag{Structural model}$$

Table 13: Structural model

Parameter	Coefficient	Standard error	z-value	p-value	Stand. coefficient
β_1	-0.161	0.033	-4.824	<0.001	-0.172
β_2	-0.376	0.024	-15.948	<0.001	-0.548
β_3	0.206	0.064	3.214	0.001	0.136
β_4	-0.442	0.056	-7.956	<0.001	-0.427

Moreover, the latent variables social functioning and depression have a residual variance, since they are not exogenous in nature. In the standardized solution, the residual variance indicates the portion of the variance that is not accounted for by the latent variable. Both depression and social functioning have a high residual variance: 0.737 and 0.690, respectively. This insight indicates that something is not entirely correct with the model, since these residual covariances should be lower if the latent variables are being predicted correctly.

Strongly related to the structural model is the notion of discriminant validity. Discriminant validity gives an indication that theoretically different constructs should not be highly intercorrelated. In other words, if two latent variables are highly correlated they could represent the same construct and they could be merged into one latent variable to obtain a more parsimonious solution (Brown, 2015). The low and insignificant ($p=0.108$) standardized covariance or correlation of -0.1 between harm avoidance and self-directedness indicates that there is little evidence for poor discriminant validity.

3.4 Goodness of fit

Fourth, the goodness of fit of the model will be evaluated using χ^2, SRMR, RMSEA, CFI and TLI. The sources of misfit will be further investigated using modification indices. It was previously concluded that there are some problems with the measurement model. The structural model seemed fine, but the results were a bit counterintuitive. This evaluation step is therefore crucial to further determine if the model is a good fit for the data or not.

3.4.1 Test statistics

The χ^2 statistic is closely related to the fit of the model and is very popular in the literature, but it has received some important criticisms. It has been noted that it is inflated by sample size and in many instances the underlying distribution is not χ^2 distributed (Brown, 2015). In this illustration, the test statistic of 505.23 is larger than the critical value of 105.52. Hence, the null hypothesis that this model is equal to a perfectly fitting model can be rejected and poor model fit is concluded. However, given the large sample size this result should not be trusted.

Absolute fit indices have therefore been employed. They assess the quality of the solution without taking into account model parsimony. First, the standardized root mean square residual (SRMR) can be interpreted as the square root average standardized residual covariance (polychoric correlation). It can be calculated using the following equation, where p is the number of indicators and ϵ is the vector of the standardized residual covariances (Shi and Maydeu-Olivares, 2020). In this illustration a SRMR of 0.081 was obtained, which indicates borderline poor model fit as it is just above the target of 0.08.

$$SRMR = \sqrt{\frac{\epsilon\epsilon}{p(p+1)/2}} \qquad (4)$$

$$RMSEA = \sqrt{\frac{\chi^2 - df}{N \times df}} \qquad (5)$$

Second, the root mean square error of approximation (RMSEA) is based on the χ^2 statistic and takes into account the error of approximation in the population. The RMSEA takes values between zero and one and the fit of the model is acceptable if it falls under 0.05. A borderline unacceptable fit is obtained with a RMSEA of 0.058.

The CFI and TLI are two comparative fit indices that will be evaluated as well. This group of statistics is called comparative, since they make a comparison between a restricted null model and an alternative model supplied by the model-builder (Brown, 2015). The comparative fit index (CFI) and Tucker-Lewis index (TLI) have been shown below. Both measures have a range of possible values from zero to one and make a correction for complexity through the degrees of freedom. Values that are close to one imply a good model fit, since the alternative and null model will then be close to each other. Generally, 0.9 is taken as a target value. In this work the CFI and TLI are, respectively, 0.952 and 0.945. To sum up, the fit of this model is borderline good or bad, depending on which fit measures are taken into account.

$$CFI = \frac{(\chi^2 - df)_{null} - (\chi^2 - df)_{alternative}}{(\chi^2 - df)_{null}} \qquad (6)$$

$$TLI = \frac{(\chi^2/df)_{null} - (\chi^2/df)_{alternative}}{(\chi^2/df)_{null}} \qquad (7)$$

Table 14: Test statistics

Statistic	**Value**	**Target**
χ^2	505.23	< 105.52
CFI	0.952	> 0.9
TLI	0.945	> 0.9
RMSEA	0.058	< 0.05
SRMR	0.081	< 0.08

3.4.2 Modification indices

The modification indices can be used to more precisely investigate sources of model misfit. They can be calculated for each fixed and constrained parameter in the model and indicate how much the model

Table 15: The 10 highest modification indices of base model

Left hand side	**Operation**	**Right hand side**	**Modification index**	**Expected parameter change**	**Stand. expected parameter change**
C1SE1BB	correlation	C1SE1D	90.619	0.329	0.473
C1SE14O	correlation	C1SE14R	44.130	0.317	0.778
social_functioning	loading	C1SE14P	36.950	-0.551	-0.311
C1SE1BB	correlation	C1SE1HH	34.474	-0.290	-0.570
social_functioning	loading	C1PA65	29.688	-0.415	-0.234
social_functioning	loading	C1PA66	26.101	-0.403	-0.227
social_functioning	loading	C1SE14O	22.440	0.463	0.261
depression	loading	C1SE1D	20.018	-0.253	-0.215
social_functioning	loading	C1SE7V	19.761	0.255	0.144
C1SE1HH	correlation	C1SE1J	19.647	0.170	0.424

χ^2 would drop if a certain parameter were to be freely estimated. A good fitting model should then also produce modification indices that are small in magnitude. A modification index that is greater than 3.84 indicates that the model fit can be significantly improved if the parameter is freely estimated (Brown, 2015). Unfortunately, the summary shown in Table 15 indicates that there are various sources of badness of fit associated with the measurement model. The question is now how to proceed. Since structural equation modeling is at heart a theory testing framework, it is important to first investigate the reason why a certain modification index is too large. A proper theoretical justification needs to be present before freeing a parameter.

First, a huge drop in the model's χ^2 of 90.62 can be realized by allowing a correlated error term between the variables 1SE1BB and SE1D, which share a loading on social functioning. On the one hand, SE1BB assesses whether the respondent believes other people would describe him/her as a giving person. On the other hand, SE1D evaluates whether the respondent believes other people see him/her as loving and affectionate. Personally, I believe that it is very plausible that these two variables are related to one another. A correlation between these variables will therefore be allowed in the improved model.

Next, the modification indices indicate that the model fit would improve dramatically should a correlated error term be allowed between the self-directedness variables 1SE14O and 1SE14R. In 1SE14O and 1SE14R it is asked whether the respondent likes to make plans for the future and knows what to want out of life, respectively. Personally, I cannot see how these variables would be directly related to each other.

Third, it is suggested that cross-loading can be allowed in the model to improve its fit. Specifically, the variable SE14P is suggested to load on social functioning. In 1SE14P it is asked whether the interviewee likes to make plans for the future. Clearly, this variable is more related to self-directedness than social functioning. Allowing such a cross-loading would also make the structural model more difficult to interpret.

Additionally, the modification indices indicate that a correlated error term should be allowed between the variables 1SE1HH and 1SE1BB. Both variables have a loading on social functioning. 1SE1HH indicates whether the respondent has experienced many warm and trusting relationships. 1SE1BB assesses whether the respondent believes he or she would be described by others as a giving person, willing to share his or her time. Personally, I cannot see how both variables are directly related to each other.

4 Expanding the base model

5 Conclusion