

Maria Seppänen

ELPYMISEN MITTARIN FAKTORIRAKENTEEN PYSYVYYS AJASSA

Informaatioteknologian ja viestinnän tiedekunta
Kandidaattitutkielma
Toukokuu 2022

Tiivistelmä

Maria Seppänen: Elpymisen mittarin faktorirakenteen pysyvyys ajassa
Kandidaattitutkielma
Tampereen yliopisto
Matematiikan ja tilastollisen data-analyysin tutkinto-ohjelma
Toukokuu 2022

Faktorianalyysi on rakenneyhtälömallien erikoistapaus ja sillä voidaan mallintaa rakenteita, joissa jatkuvilla havaituilla muuttujilla mitataan faktoreita eli latenteja muuttujia, joita ei voida suoraan mitata tai havaita. Faktoreilla pyritään selittämään havaittujen muuttujien välisiä korrelaatioita. Faktorien välisiä yhteyksiä voidaan mallintaa toisen kertaluvun faktorien avulla. Tämän tutkielman tarkoituksena on tutkia elpymisen mittarille luodun kolmen faktorin rakenteen pysyvyyttä ajassa. Tutkittava ROS-mittari (Restoration Outcome Scale) pohjautuu aikaisempiin elpymiseen liittyviin tutkimustuloksiin.

Aineistona toimii kolmena peräkkäisenä vuonna toteutettu sähköinen työntekijäkysely. Tutkimuksessa tutkittiin luonnolle altistumisen vaikutusta työntekijöiden hyvinvointiin. ROS-mittarin faktorirakenne muodostuu kolmesta faktorista: keskittymiskyvyn elpyminen, rentoutuminen ja rauhallisuus, ajatusten kirkastuminen. Niiden ajatellaan olevan mitattavissa kyselyn kuudella kysymyksellä, jotka muodostavat tarvittavat kuusi havaittua muuttujaa. Nämä kuusi havaittua muuttujaa mittaavat faktoreita niin, että kolme muuttujaa latautuvat rentoutumiselle ja rauhallisuudelle, yksi keskittymiskyvyn elpymiselle ja kaksi ajatusten kirkastumiselle.

Elpymisen mittarin faktorirakenne mallinnetaan erikseen jokaiselle faktorille kolmessa aikapisteessä niin, että kullekin faktorimallille luodaan toisen kertaluvun faktori, jota hyödynnetään rakenteen pysyvyyden tutkimiseksi yli ajan. Uskottavuussuhdetestillä verrataan, sopiiko aineistoon paremmin malli, joka olettaa faktorirakenteen samaksi jokaisessa aikapisteessä vai malli, jossa rakenteen annetaan muuttua ajassa.

Tutkielman tulos on, että erikseen tarkasteltuna kaikkien kolmen ROS-mittarin faktorien faktorirakenne on pysyvä ajassa. Kukin kolmesta uskottavuussuhdetestistä johti tulokseen, että faktorimalli, jonka faktorirakenne oletettu samaksi yli ajan, sopii aineistoon paremmin kuin malli, jonka rakenteen annetaan muuttua ajassa. Tämä viittaa siihen, että ROS-mittarin faktorirakenne on pysyvä ajassa.

Avainsanat: faktorianalyysi, uskottavuussuhdetesti, pitkittäisaineisto

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

Sisällys

1	Johdanto	4
2	Tutkimusaineisto	5
2.1	Elpymismuuttujien keskiarvot ja korrelaatiot	5
3	Faktorianalyysin teoriaa	8
3.1	Faktorianalyysi	8
3.2	Konfirmatorisen faktorimallin spesifointi	8
3.3	Teoreettinen kovarianssirakenne	9
3.4	Graafiset esitykset	10
3.5	Mallin identifioituvuus	12
3.6	Toisen kertaluvun faktorimalli	12
4	Suurimman uskottavuuden menetelmä	14
4.1	Hypoteesien testaaminen	14
5	Faktorianalyysin sovellusesimerkki	16
5.1	Hypoteettinen faktorirakenne	16
5.2	Mallin luominen hypoteettisen rakenteen pohjalta	16
5.2.1	Ensimmäisen faktorin ensimmäisen asteen malli	16
5.2.2	Toisen faktorin ensimmäisen asteen malli	17
5.2.3	Kolmannen faktorin ensimmäisen asteen malli	17
5.3	Toisen kertaluvun faktorimalli hypoteettisen rakenteen pohjalta	18
5.3.1	Mallit	18
5.4	Rakenteen testaaminen	21
6	Yhteenveto	27
	Lähteet	28

1 Johdanto

Faktoriansalyysillä on pitkä historia psykologiassa jo 1900-luvun alkupuolelta. Faktoriansalyysin isänä pidetty psykologi Charles Spearman esitti vuonna 1904 ensimmäisenä mallin, jossa monen havaitun muuttujan avulla mitattiin yhtä faktoria – älykkyyttä. Myöhemmin 1940-luvulla suurimman uskottavuuden menetelmää ja myös hypoteesien testausta sovellettiin faktorimalleihin. Faktoriansalyysin käyttö yleistyi kuitenkin vasta tietokoneiden käytön yleistyessä, sillä ennen sitä se oli laskennallisesti hyvin työlästä. Konfirmatorisen faktoriansalyysin juuret ovat 1960-luvulla, kun alettiin kiinnostua latausmatriisille asetetuista rajoituksista (Leskinen 1987.)

Nykypäivänä faktoriansalyysiä voi toteuttaa monella eri ohjelmapaketilla eri ohjelmistoissa, kuten SAS, SPSS sekä R. Tämä tutkielma käsittelee konfirmatorista faktoriansalyysiä ja faktorirakenteen pysyvyyttä, kun käytössä on pitkittäisaineisto ja laskemiseen käytetään R-ohjelmiston *lavaan*-pakettia. Tutkielman kannalta tärkeää kirjallisuutta on ollut erityisesti *Structural Equations with Latent Variables* (Bollen 1989) sekä *Faktoriansalyysi* (Leskinen 1987).

Tutkielma alkaa tutkimusaineiston esittelyllä luvussa kaksi, jossa kerrotaan sen alkuperä ja sovellusesimerkissä käytetyt muuttujat. Kolmannessa luvussa käydään läpi faktoriansalyysin perusteoriaa, kuten yleinen malli sekä mallin identifioituvuus ja toisen kertaluvun faktorimalli. Neljännessä luvussa tarkastellaan suurimman uskottavuuden menetelmän soveltamista faktorilatauksien yhtäsuuruuden tutkimiseen. Viidennessä luvussa edellä käsiteltyä teoriaa sovelletaan käytännön aineistoon ja tarkastellaan aineiston hypoteettisen faktorirakenteen pysyvyyttä ajassa.

2 Tutkimusaineisto

Tutkimusaineistona toimii kolmena peräkkäisenä vuonna toteutetun työntekijäkyselyn vastaukset. (Korpela, Nummi et al. 2017) Kyselyllä pyrittiin selvittämään luonnolle altistumisen vaikutusta työntekijöiden hyvinvointiin. Luonnolle altistumista mitattiin itseilmoittamana sekä työpaikalla, että kotona. Aineistossa on kolme aikapistettä. Otokoko on 664 työntekijää.

Tähän tutkimukseen käytetään ainoastaan kuutta muuttujaa, joista muodostuu elpymistä mittaava ROS-mittari (Restoration Outcome Scale). Mittari luotiin aikaisempien toimenpiteiden sekä tulosten pohjalta liittyen elpymiseen. (Korpela, Ylén et al. 2008). Siihen sisältyy kuusi kysymystä, joihin osallistujat vastaavat 7-portaisella Likert-asteikolla välillä “ei ollenkaan” ja “täysin”. Kysymykset muotoiltiin seuraavasti “Arvioi, missä määrin seuraavat väittämät kuvaavat vapaa-ajan toimintojasi? Vapaa-ajan toimintoni...”

- ... rauhoittavat minua.
- ... lisäävä keskittymiskykyäni ja valppauttani.
- ... antavat uutta intoa ja pirteyttä arkiaskareisiini.
- ... saavat minut tuntemaan itseni elpyneeksi ja rentoutuneeksi.
- ... saavat minut unohtamaan kaikki arkipäivän huolet ja murheet.
- ... auttavat minua selkeyttämään ja kirkastamaan ajatuksiani.

Tässä tutkielmassa havaittuihin muuttujiin tullaan jatkossa viittaamaan lyhyemmin muodossa “elpyminen1”, “elpyminen2”, ... ,“elpyminen 6”, samassa järjestyksessä kuin yllä.

Aineistosta löytyy 25 havaintoa, joissa vastaaja ei ole vastannut jokaisena vuonna, tai dataa puuttuu yksittäisistä vastauksista. Sillä tällaisia havaintoja on pieni määrä suhteessa otoskokoon ja sen oletetaan olevan satunnaista, ennen faktorimallin luomista kaikki puuttuvaa dataa sisältävät havainnot poistetaan. Aineiston 25 puuttuvaa tietoa sisältävät havainnot poistetaan, eli 664 havainnosta jäljelle jää poistamisen jälkeen 639, joita käytetään luvussa 5 faktorimallien luomiseen.

2.1 Elpymismuuttujien keskiarvot ja korrelaatiot

Tarkastellaan aineiston muuttujien tunnuslukuja. Jokaisen havaitun muuttujan mediaani on 5, eli jokaisen kysymyksen vastaukset ovat painottuneet keskelle. Taulukossa 2.1 ovat muuttujien keskiarvot, joista saadaan enemmän vastausten vaihtelua esille. Pienimmät keskiarvot ovat muuttujilla elpyminen5 ja elpyminen6. Suurimmat keskiarvot taas muuttujalla elpyminen1. Keskiarvot ovat samalla muuttujalla lähellä toisiaan tai jopa samoja eri mittauskerroilla.

Taulukko 2.1. Aineiston muuttujien keskiarvot.

Muuttuja	Ensimmäinen mittauskerta	Toinen mittauskerta	Kolmas mittauskerta
Elpyminen1	5,31	5,34	5,29
Elpyminen2	5,08	5,13	5,07
Elpyminen3	5,23	5,23	5,24
Elpyminen4	5,17	5,17	5,16
Elpyminen5	4,82	4,80	4,67
Elpyminen6	4,97	5,00	4,98

Muuttujien välisiä yhteyksiä voidaan tarkastella korrelaatioiden avulla. Taulukoista 2.2 – 2.4 löytyvät muuttujien väliset korrelaatiot jokaisella mittauskerralla. Korrelaatio on kaikkien muuttujien välillä positiivista ja yli 0,50. Korrelaatiot ovat hyvin lähellä toisiaan mittauskerroista riippumatta, niiden keskinäinen ero on aina alle 0,09 yksikköä.

Taulukko 2.2. Aineiston muuttujien väliset korrelaatiot ensimmäisellä mittauskerralla.

	Elpyminen1	Elpyminen2	Elpyminen3	Elpyminen4	Elpyminen5	Elpyminen6
Elpyminen1	1					
Elpyminen2	0,74	1				
Elpyminen3	0,73	0,78	1			
Elpyminen4	0,72	0,74	0,81	1		
Elpyminen5	0,57	0,58	0,63	0,69	1	
Elpyminen6	0,62	0,75	0,74	0,75	0,71	1

Taulukko 2.3. Aineiston muuttujien väliset korrelaatiot toisella mittauskerralla.

	Elpyminen1	Elpyminen2	Elpyminen3	Elpyminen4	Elpyminen5	Elpyminen6
Elpyminen1	1					
Elpyminen2	0,74	1				
Elpyminen3	0,69	0,76	1			
Elpyminen4	0,66	0,71	0,80	1		
Elpyminen5	0,53	0,53	0,60	0,69	1	
Elpyminen6	0,57	0,69	0,69	0,72	0,66	1

Taulukko 2.4. Aineiston muuttujien väliset korrelaatiot kolmannella mittauskerralla.

	Elpyminen1	Elpyminen2	Elpyminen3	Elpyminen4	Elpyminen5	Elpyminen6
Elpyminen1	1					
Elpyminen2	0,80	1				
Elpyminen3	0,73	0,80	1			
Elpyminen4	0,74	0,77	0,81	1		
Elpyminen5	0,54	0,57	0,59	0,67	1	
Elpyminen6	0,61	0,67	0,70	0,74	0,70	1

Pienin korrelaatio on havaittavissa muuttujan elpyminen5 ja muuttujien elpyminen1 sekä elpyminen2 välillä. Näiden muuttujien pareittaiset korrelaatiot ovat jokaisella mittauskerralla alhaisimmat, 0,53-0,58 välillä. Korkein korrelaatio jokaisella mittauskerralla on muuttujien elpyminen4 ja elpyminen3 välillä, minkä arvo on 0,80-0,81 mittauskerrasta riippuen. Valtaosa muuttujista ovat keskenään korreloituneita niin, että korrelaatio on jotain 0,60-0,75 välillä.

3 Faktorianalyysin teoriaa

3.1 Faktorianalyysi

Faktorianalyysillä pyritään selittämään havaittujen muuttujien välisiä yhteyksiä havaitsemattomilla piilomuuttujilla, eli latenteilla muuttujilla. Havaittujen muuttujien taustalla ajatellaan olevan jokin ei suoraan mitattavissa oleva ominaisuus tai ilmiö, joka aiheuttaa muuttujien välistä korrelaatiota. Faktorien eli kyseessä olevien latenttien muuttujien on tarkoitus selittää tätä aineistossa ilmenevää korrelaatiota. Muodostamalla jokaiselle havaitulle muuttujalle eli indikaattorimuuttujalle oma regressioyhtälönsä, luodaan indikaattorimuuttujien ja faktorien välille yhteys, joka sitten myös muodostaa mallin kovarianssirakenteelle.

Faktorianalyysi voidaan karkeasti sanottuna jakaa kahteen eri osa-alueeseen: eksploratiiviseen ja konfirmatoriseen faktorianalyysiin. Eksploratiivisessa faktorianalyysissä malli pyritään luomaan aineiston pohjalta, kun taas konfirmatorisessa faktorianalyysissä oletetaan etukäteen faktorien määrä, korrelaatiot ja muut mahdolliset rajoitteet. Tällöin mallin sopivuutta voidaan testata erilaisin tilastollisin menetelmin ja tarvittaessa muuttaa alkuoletuksia, jotta mallista tulisi sopivampi. Tässä tutkielmassa keskitytään konfirmatoriseen faktorianalyysiin.

3.2 Konfirmatorisen faktorimallin spesifiointi

Faktorimallin yleinen muoto on

$$(3.1) \quad \mathbf{y} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\epsilon}$$

missä \mathbf{y} on havaittu muuttuja, $\boldsymbol{\eta}$ on latentti muuttuja ja $\boldsymbol{\epsilon}$ mittausvirhe. Matriisi Λ_y sisältää tiedon latentin muuttujan vaikutuksesta havaitulle muuttujalle.

Latausmatriisi on $m \times n$ matriisi, jossa m on indikaattorimuuttujien lukumäärä ja n on latenttien muuttujien lukumäärä. Jokainen matriisin sarake kuvastaa yhtä latenttia muuttujaa, jokainen rivi yhden muuttujan latautumista eri faktoreille. Käytännössä, lataus tarkoittaa havaitun muuttujan yksikön odotettua muutosta, kun latentti muuttuja muuttuu yhden yksikön verran.

Faktorimallin lähtöoletuksia on, että faktorien ja mittausvirheiden välinen korrelaatio on nolla, mittausvirheet eivät korreloi keskenään ja niiden odotusarvo on nolla. (Bollen 1989, s. 233–234) Tämä voidaan merkitä

$$\begin{aligned} \text{Cov}(\eta_i, \epsilon_i) &= 0, \\ \text{Cov}(\epsilon_i, \epsilon_j) &= 0, \\ E(\epsilon_i) &= 0, \end{aligned}$$

missä i ja j ovat faktorien indeksejä, kun $i \neq j$.

Esimerkki 3.1. Otetaan esimerkiksi tilanne, jossa on neljä havaittua muuttujaa. Ajatellaan, että havaittujen muuttujien taustalla on yksi piilomuuttuja. Yhtälön 3.1 voi silloin merkitä

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} [\eta_1] + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix},$$

missä latenttien muuttujien lataukset havaitulle muuttujalle ovat λ_{11} , λ_{21} , λ_{31} ja λ_{41} . Tämän voi merkitä jokaiselle havaitulle muuttujalle erikseen

$$\begin{cases} y_1 = \lambda_{11}\eta_1 + \epsilon_1 \\ y_2 = \lambda_{21}\eta_1 + \epsilon_2 \\ y_3 = \lambda_{31}\eta_1 + \epsilon_3 \\ y_4 = \lambda_{41}\eta_1 + \epsilon_4 \end{cases}.$$

Oletetaan, että

$$\begin{aligned} Cov(\eta_i, \epsilon_i) &= 0, \\ E(\epsilon_i) &= E(\eta_i) = 0, \text{ jossa } i \text{ faktorin indeksi.} \end{aligned}$$

Tässä tapauksessa havaittu muuttuja mittaa vain yhtä faktoria. Tämä ilmenee latausmatriisissa niin, että jokaisella rivillä on yksi nollasta poikkeava lataus. Jos taustamuuttujia olisi enemmän kuin yksi, olisi jokaisella rivillä yksi nollasta poikkeava lataus ja muut arvot olisivat nollia.

3.3 Teorettinen kovarianssirakenne

Merkitään faktorien λ kovarianssimatriisia Φ :llä ja mittavirheiden ϵ kovarianssia Θ :lla. Tällöin havaittujen muuttujien teorettinen kovarianssimatriisi Σ on

$$\begin{aligned} \Sigma &= E[\mathbf{y}\mathbf{y}'] \\ &= E[(\Lambda_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta}'\Lambda_y' + \boldsymbol{\epsilon}')] \\ (3.2) \quad &= \Lambda_y E[\boldsymbol{\lambda}\boldsymbol{\lambda}']\Lambda_y' + E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] \\ &= \Lambda_y Cov(\boldsymbol{\lambda})\Lambda_y' + Cov(\boldsymbol{\epsilon}) \\ &= \Lambda_y\Phi\Lambda_y' + \Theta_\epsilon. \end{aligned}$$

Esimerkki 3.2. Tarkastellaan samaa rakennetta kuin esimerkissä 3.1, eli yhden faktorin ja neljän havaitun muuttujan faktorimallia 3.1, joka toteuttaa faktorimallin edellyttämät ehdot. Tällöin kovarianssimatriisin laskemiseen tarvitaan latausmatriisi Λ_y , faktorien kovarianssimatriisi Φ sekä mittausvirheiden kovarianssi Θ_ϵ .

$$\Lambda_y = [\lambda_{11} \quad \lambda_{21} \quad \lambda_{31} \quad \lambda_{41}]^T$$

$$\Phi = [\phi_{11}]$$

$$\Theta_\epsilon = \begin{bmatrix} \text{Var}(\epsilon_1) & & & \\ 0 & \text{Var}(\epsilon_2) & & \\ 0 & 0 & \text{Var}(\epsilon_3) & \\ 0 & 0 & 0 & \text{Var}(\epsilon_4) \end{bmatrix}$$

Käyttäen avuksi mallin tietoja, voidaan kovarianssimatriisin 3.2 rakenteen pohjalta kirjoittaa:

$$\Sigma = \begin{bmatrix} \lambda_{11}^2 \phi_{11} + \text{Var}(\epsilon_1) & & & \\ \lambda_{21} \lambda_{11} \phi_{11} & \lambda_{21}^2 \phi_{11} + \text{Var}(\epsilon_2) & & \\ \lambda_{31} \lambda_{11} \phi_{11} & \lambda_{31} \lambda_{21} \phi_{22} & \lambda_{31}^2 \phi_{11} + \text{Var}(\epsilon_3) & \\ \lambda_{41} \lambda_{11} \phi_{11} & \lambda_{41} \lambda_{21} \phi_{11} & \lambda_{41}^2 \lambda_{31} \phi_{11} & \lambda_{41}^2 \phi_{11} + \text{Var}(\epsilon_4) \end{bmatrix},$$

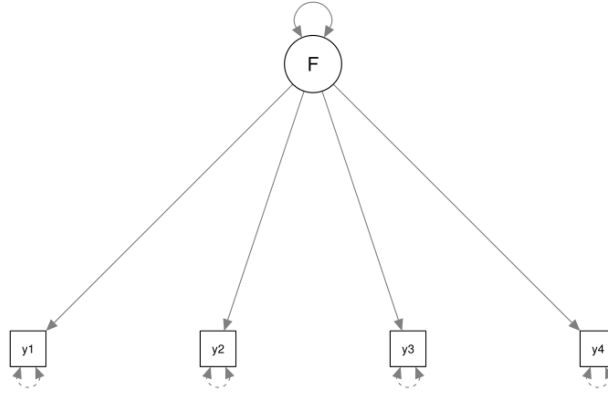
koska matriisi on symmetrinen, on tässä raportoitu vain diagonaali ja sen alapuoliset alkiot. Tämä matriisi on siis oletetusta mallista muodostunut teoreettinen kovarianssirakenne. Käytössä olevan aineiston pohjalta voi laskea myös havaittu kovarianssimatriisi S , joka voidaan merkitä havaittujen muuttujien varianssien ja kovarianssien avulla

$$S = \begin{bmatrix} \text{Var}(y_1) & & & \\ \text{Cov}(y_2, y_1) & \text{Var}(y_2) & & \\ \text{Cov}(y_3, y_1) & \text{Cov}(y_3, y_2) & \text{Var}(y_3) & \\ \text{Cov}(y_4, y_1) & \text{Cov}(y_4, y_2) & \text{Cov}(y_4, y_3) & \text{Var}(y_4) \end{bmatrix}.$$

3.4 Graafiset esitykset

Rakenneyhtälömalleja ja faktorimalleja on tapana visualisoida polkugraafeilla, joista voi nähdä muuttujien välisiä yhteyksiä. Tässä tutkielmassa polkugraafeissa neliö kuvastaa havaittua muuttujaa eli indikaattorimuuttujaa ja ympyrä latenttia muuttujaa. Nuolet muuttujista toisiin kuvaavat muuttujan latautumista toiselle muuttujalle, kun taas nuoli muuttujasta itseensä on kyseisen muuttujan varianssi. Kaksisuuntainen nuoli merkitsee kovarianssia. Katkoviivanuoli tarkoittaa, että parametri on kiinnitetty joksikin arvoksi. Kiinteiden nuolien kohdalle on merkitty kyseisen parametrin estimaatti, katkoviivanuolen tapauksessa näkyy parametrille asetettu arvo.

Käytetään mallien luomiseen R-ohjelmistoa ja sen lavaan-kirjastosta löytyvää *cfa()*-funktioita. Mallien piirtämiseen käytetään kirjastoa *semPlot* ja sieltä löytyvää *semPaths()*-funktioita. Tulevissa esimerkeissä estimaattien arvoja ei näy graafisissa esityksissä, mutta ne näkyvät tutkielman loppupuolella, kun tarkastellaan sovellusesimerkin faktorimalleja luvussa 5.



Kuva 3.1. Yhden faktorin ja neljän indikaattorimuuttujan polkukaavio.

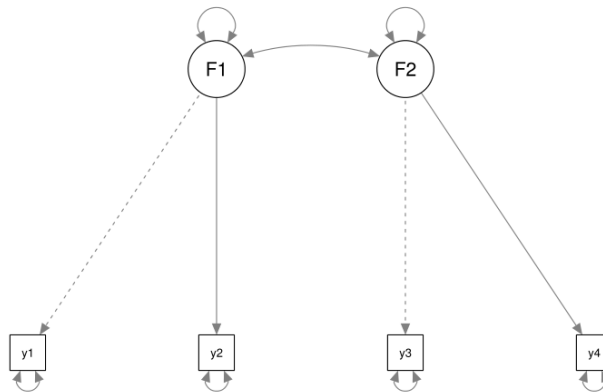
Esimerkki 3.3. Tarkastellaan graafista esitystä esimerkille 3.1. Mallissa neljä indikaattorimuuttujaa y_1, y_2, y_3 ja y_4 latautuvat yhdelle latentille muuttujalle F . Asetetaan indikaattorimuuttujien varianssit arvoksi yksi. Tämän mallin graafinen esitys löytyy kuvasta 3.1.

Esimerkki 3.4. Tarkastellaan polkugraafia tilanteessa, jossa latenteja muuttujia on kaksi ja indikaattorimuuttujia neljä, joista kaksi latautuvat kullekin latentille muuttujalle. Mallin yhtälö olisi silloin seuraavanlainen:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix},$$

missä y_1, \dots, y_4 ovat indikaattorimuuttujia, $\lambda_{11}, \lambda_{21}$ ensimmäisen faktorin latauksia, $\lambda_{32}, \lambda_{42}$ toisen faktorin latauksia, η_1, η_2 faktoreita ja $\epsilon_1, \dots, \epsilon_4$ mittavirheitä.

Mallissa on siis neljä indikaattorimuuttujaa y_1, y_2, y_3 ja y_4 ja kaksi latenttia muuttujaa F_1 ja F_2 . Kyseisessä mallissa kunkin latentin muuttujan yhden indikaattorimuuttujan lataus on kiinnitetty arvoksi 1, mikä näkyy kuviossa katkoviiavana. Mallin graafinen esitys löytyy kuvasta 3.2.



Kuva 3.2. Kahden faktorin ja neljän indikaattorimuuttujan polkukaavio, missä y_1 :n sekä y_3 :n latauksen arvo on kiinnitetty.

3.5 Mallin identifioituvuus

Mallin identifioituvuus on tärkeä kysymys rakenneyhtälömalleja tarkastellessa. Tässä tutkielmassa siihen keskitytään kuitenkin vain konfirmatorisen faktorianalyysin näkökulmasta. Faktorimallissa on usein monta estimoitavaa parametria, joten herää kysymys - onko jokaiselle parametrille olemassa uniikki ratkaisu? Tähän kysymyksen pyritään vastaamaan tarkastelemalla mallin identifioituvuutta.

Ali- ja yli-identifioituvuus

Jos jokaiselle parametrille löytyy uniikki ratkaisu, sanotaan mallin olevan identifioituva, jos taas ei, malli on ali-identifioituva. Jokaiselle tuntemattomalle estimoitavalle parametrille pyritään löytämään yhtälö tunnettujen parametrien avulla. Jos tällainen on olemassa jokaiselle parametrille ja yhtälöt johtavat yksiselitteisiin tuloksiin, ovat parametrit identifioituvia. Muussa tapauksessa yhdelle tai useammalle parametrille ei löydy yksiselitteistä tulosta, jolloin parametreista muodostuva mallikaan ei ole identifioituva tai yksiselitteinen. Tällaista mallia kutsutaan ali-identifioituvaksi.

Ali-identifioituvasta mallista saadaan identifioituva asettamalla rajoitteita parametrien arvoille, kuten asettaa ne yhdeksi tai nolaksi tai asettamalla yhtäsuuruuksia parametrien välille. Esimerkiksi, jos mallin mittausvirheiden ajatellaan olevan korreloitumattomia, niin mittausvirheiden kovarianssimatriisiin jää nolasta poikkeavaksi vain diagonaali. Parametreille on mahdollista asettaa yhtäsuuruuksia, joka vähentää estimoitavien parametrien lukumäärää. Rajoitteiden asettaminen on oleellinen osa latenttien muuttujien skaalan asettamista. Yleisimmät tavat ovat asettaa latenttien muuttujien varianssit, tai yksi jokaisen latentin muuttujan latauksista arvoksi yksi.

Parametri voi olla yli-identifioituva, jos yksi parametri voi saada enemmän kuin yhden yhtälön tunnettujen parametrien avulla merkittynä. Jos mallissa on yksi tai enemmän yli-identifioituvaa parametria, myös malli on yli-identifioituva. Tavoite on, että ali-identifioituvasta mallista saadaan joko identifioituva, jolloin jokainen tuntematon parametri on identifioituva tai yli-identifioituva.

3.6 Toisen kertaluvun faktorimalli

Jos ensimmäisen kertaluvun faktoreilla pystytään mallintamaan indikaattorimuuttujien korrelaatiota, niin tämän tehdä myös itse faktoreille. Jos faktorit korreloivat keskenään, voidaan niiden taustalla ajatella olevan jokin latentti muuttuja, joka selittää faktorien korrelaation. Tällainen muuttuja on toisen kertaluvun faktori (Jöreskog ja Sörbom 1979, s. 65–68). Merkitään toisen kertaluvun faktorirakennetta

$$\eta = \Gamma \xi + \zeta,$$

jossa Γ on latausmatriisi, ξ sisältää toisen kertaluvun faktorit ja ζ on mittausvirhevektori.

Toisen kertaluvun faktorirakenteen avulla voidaan tehdä kovarianssimatriisiesitys ensimmäisen kertaluvun faktoreille. Tämä kovarianssirakenne on hyvin samankaltai-

nen ensimmäisen kertaluvun faktorimallin havaitun muuttujan kovarianssirakenteen kanssa. Rakenne voidaan muodostaa seuraavasti:

$$\begin{aligned}
 \text{Cov}(\boldsymbol{\eta}) &= \boldsymbol{\Phi} = \mathbf{E}[\boldsymbol{\eta}\boldsymbol{\eta}'] \\
 &= \mathbf{E}[(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})(\boldsymbol{\xi}'\boldsymbol{\Gamma}' + \boldsymbol{\zeta}')] \\
 &= \boldsymbol{\Gamma}\mathbf{E}[\boldsymbol{\xi}\boldsymbol{\xi}']\boldsymbol{\Gamma}' + \mathbf{E}[\boldsymbol{\zeta}\boldsymbol{\zeta}'] \\
 &= \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}' + \boldsymbol{\Psi},
 \end{aligned}$$

jossa $\boldsymbol{\Phi}$ on ensimmäisen kertaluvun faktorien kovarianssimatriisi, $\boldsymbol{\eta}$ ensimmäisen kertaluvun faktori, $\boldsymbol{\Gamma}$ latausmatriisi, $\boldsymbol{\xi}$ toisen kertaluvun faktori, $\boldsymbol{\Omega}$ toisen kertaluvun faktorien kovarianssimatriisi, $\boldsymbol{\zeta}$ mittausvirhe ja $\boldsymbol{\Psi}$ mittausvirhematriisi.

Toisen kertaluvun faktorimalli kokonaisuudessaan on muotoa

$$(3.3) \quad \begin{cases} \mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon} \\ \boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \end{cases},$$

jossa $\boldsymbol{\Lambda}$ on ensimmäisen kertaluvun faktorien latausmatriisi, $\boldsymbol{\eta}$ ensimmäisen kertaluvun faktorit, $\boldsymbol{\Gamma}$ toisen kertaluvun faktorien latausmatriisi, $\boldsymbol{\xi}$ toisen kertaluvun faktorit, $\boldsymbol{\epsilon}$ ensimmäisen kertaluvun mittausvirhematriisi, $\boldsymbol{\zeta}$ toisen kertaluvun mittausvirhematriisi (Leskinen 1987, 156–158.)

Tällöin toisen kertaluvun kovarianssimatriisiesitys voidaan asettaa ensimmäisen kertaluvun kovarianssimatriisiesitykseen seuraavasti:

$$\begin{aligned}
 \boldsymbol{\Sigma} &= \boldsymbol{\Lambda}_y\boldsymbol{\Phi}\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_\epsilon \\
 &= \boldsymbol{\Lambda}[\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}]\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_\epsilon.
 \end{aligned}$$

Toisen kertaluvun faktorimallin identifioitavuus voidaan tarkistaa samalla periaatteella kuin ensimmäisen asteen faktorimallinkin eli tuntemattomat parametrit on voitava estimoida käyttäen tunnettuja parametreja.

4 Suurimman uskottavuuden menetelmä

Faktoriansalyysissä estimoidaan muodostuneen kovarianssirakenteen parametrit, jotka voidaan tehdä suurimman uskottavuuden menetelmällä. Tällöin estimaateilla on suurella otoskoolla hyvät ominaisuudet, kun \mathbf{y} on multinomijakautunut. (Jöreskog ja Sörbom 1979, s. 19) Oletetaan, että aineiston havaitut muuttujat ovat n -kokoinen otos p -ulotteisesta normaalijakaumasta $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Otoksen $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ uskottavuusfunktio on muotoa

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{y}_1, \dots, \mathbf{y}_n) &= \prod_{i=1}^n f(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= (2\pi)^{\frac{-pn}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp -\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}^T) \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}). \end{aligned}$$

Logaritmoimalla tämä, saadaan

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\left(\frac{pn}{2}\right) \ln 2\pi - \left(\frac{n}{2}\right) \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}^T) \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}),$$

joka voidaan sieventää muotoon

$$(4.1) \quad \ln L(\boldsymbol{\Sigma}) = -\frac{n-1}{2} (\ln |\boldsymbol{\Sigma}| - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S})),$$

jossa $\boldsymbol{\Sigma}$ on estimoitavien parametrien funktiona saatu kovarianssirakenne ja \mathbf{S} otoskovarianssimatriisi.

Tyypillisesti, tämä logaritmoitu uskottavuusfunktio maksimoitaisiin, mutta käytännössä on järkevämpi tarkastella minimoitavaa funktiota F_{ML}

$$F(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Theta})_{ML} = \ln |\boldsymbol{\Sigma}| - \ln |\mathbf{S}| + \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) - p,$$

jossa $\boldsymbol{\Sigma}$ on teoreettinen kovarianssimatriisi, \mathbf{S} on otoskovarianssimatriisi ja p on \mathbf{y} :n dimensio. Tälle pätee $F_{ML} \geq 0$ (Leskinen 1987, s. 92–95).

4.1 Hypoteesien testaaminen

Parametrien estimointi suurimman uskottavuuden menetelmällä mahdollistaa suurimman uskottavuuden menetelmän käyttämisen tilastollisten hypoteesien testaamiseen, jolloin uskottavuussuhdetestillä voidaan testata esimerkiksi mallin sopivuutta aineistoon.

Jotta pystytään tutkimaan faktorirakenteen pysyvyyttä, tarkastellaan tiettyjä latauksia. Luodaan malli, jossa nämä lataukset estimoituvat vapaasti ja malli, jossa ne on rajoitettu yhtä suuriksi. Verrataan malleja uskottavuussuhdetestillä. Valitaan nollahypoteesiksi, että valittu rajoitettu faktorimalli on sopiva, ja vaihtoehtoiseksi hypoteesiksi, että vapaammin estimoitava malli on sopiva. Nämä voidaan kirjoittaa seuraavasti:

$$\begin{aligned}
H_0 : \Sigma &= \Lambda_* \Omega_* \Lambda_*^T + \Theta_*, \\
H_1 : \Sigma &= \Lambda \Omega \Lambda^T + \Theta,
\end{aligned}$$

missä alaindeksi * viittaa malliin, jossa lataukset on rajoitettu yhtä suuriksi.

Suurimman uskottavuuden periaatteen mukaan saadaan H_1 :n ollessa voimassa funktion 4.1 maksimi, kun

$$\hat{\Sigma} = \hat{\Lambda} \hat{\Omega} \hat{\Lambda}^T + \hat{\Theta},$$

ja H_0 :n ollessa voimassa, kun

$$\hat{\Sigma} = \hat{\Lambda}_* \hat{\Omega}_* \hat{\Lambda}_*^T + \hat{\Theta}_*.$$

Hypoteesien testaamiseen tarvitaan testisuure, jonka avulla voidaan arvioida hypoteeseja. Tiedetään, että H_0 :n ollessa voimassa, noudattaa uskottavuussuhteen testisuure $-2\ln(\lambda)$ asympotoottisesti χ^2 -jakaumaa. Vapausasteet ovat tällöin $t_1 - t_0$, jotka ovat kunkin hypoteesin määrittämän faktorimallin estimoitavien parametrien lukumäärä. Testisuureen funktioksi saadaan

$$\begin{aligned}
-2\ln(\lambda) &= -2\ln\left(\frac{\max_{H_0} L}{\max L}\right) \\
&= -2\ln(\max_{H_0} L) + 2\ln(\max L) \\
&= (n-1)[\ln|\hat{\Sigma}| + \text{tr}(|\hat{\Sigma}|^{-1}S) - \ln|S| - p] \\
&= (n-1)F(\hat{\Lambda}, \hat{\Phi}, \hat{\Theta})_{ML}.
\end{aligned}$$

χ^2 -testisuureen arvon laskeminen voidaan tehdä R-ohjelmiston avulla. Tällöin saadaan laskettua myös mallien vapausasteet sekä χ^2 -testisuureta vastaava p -arvo. Tähän voidaan käyttää `lrtest`-kirjastoa, sekä `lrtest()`-funktiota. Testaamisesta R-ohjelmiston avulla enemmän luvussa 5.

5 Faktorianalyysin sovellusesimerkki

5.1 Hypoteettinen faktorirakenne

Tähän analyysiin käytettävässä aineistossa on kuusi havaittua muuttujaa, kukin mitattu kolmessa aikapisteessä. Nämä muuttujat muodostavat ROS-mittarin, ja siihen pohjautuvan aikaisemman teorian pohjalta (Korpela, Ylén et al. 2008, s. 641) on luotu hypoteettinen rakenne. Havaittujen muuttujien taustalla ajatellaan olevan kolme ei-havaittavissa olevaa ilmiötä:

- rentoutuminen ja rauhallisuus,
- keskittymiskyvyn elpyminen,
- ajatusten kirkastuminen,

missä muuttujat “elpyminen1”, “elpyminen3”, “elpyminen4” mittaavat rentoutumista sekä rauhallisuutta, “elpyminen2” mittaa keskittymiskyvyn elpymistä ja “elpyminen5”, “elpyminen6” mittaavat ajatusten kirkastumista.

5.2 Mallin luominen hypoteettisen rakenteen pohjalta

Tässä tutkielmassa tarkastellaan faktorien rakennetta ajassa. Muodostetaan kolme mallia, yksi jokaiselle faktorille. Jokainen malli on rakennettu niin, että yksi malli kuvaa yhtä aikaisempaan teoriaan pohjautuvaa faktoria jokaisessa kolmessa aikapisteessä. Rakennetaan ensiksi jokaiselle teoreettiselle faktorille ensimmäisen kertaluvun malli. Laajennetaan jokainen malli toisen kertaluvun malliksi, josta tarkastellaan yläfaktoreiden latautumista aikafaktoreille. Lisäksi luodaan rajoitetumpi malli, joka olettaa yhtäsuuruuden yläfaktorien latauksille. Koska jokaista teoriaan pohjautuvaa faktoria tarkastellaan erikseen, oletetaan, että niillä ei ole keskinäistä korrelaatiota.

5.2.1 Ensimmäisen faktorin ensimmäisen asteen malli

Ensimmäisen hypoteettisen faktorin ajatellaan mittaavan rentoutumista ja rauhallisuutta. Luodaan ensimmäisen kertaluvun faktorimalli niin, että jokainen aikapiste on latentti muuttuja, jolle latautuu sen aikapisteen indikaattorimuuttujat y_1, y_4, y_7 (elpyminen1), y_2, y_5, y_8 (elpyminen3), ja y_3, y_6, y_9 (elpyminen4). Kiinnitetään lataukset $\lambda_{11} = \lambda_{12} = \lambda_{13} = 1$ skaalan asettamiseksi.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{23} \\ 0 & 0 & \lambda_{33} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix},$$

missä y_1, y_2, \dots, y_9 ovat havaitut muuttujat, λ_{ij} lataukset faktorille, jossa $i = 1, 2, 3$ hypoteettinen faktori ja $j = 1, 2, 3$ mittauskerta, $\epsilon_1, \epsilon_2, \dots, \epsilon_9$ ovat kunkin indikaattorin mittausvirheet.

5.2.2 Toisen faktorin ensimmäisen asteen malli

Toisen hypoteettisen faktorin ajatellaan mittaavan keskittymiskyvyn elpymistä. Luodaan ensimmäisen kertaluvun faktorimalli toiselle faktorille niin, että jokainen aikapiste on latentti muuttuja, jolle latautuu sen aikapisteen indikaattorimuuttuja y_1, y_2, y_3 (elpyminen2). Skaalan asettamiseksi, kiinnitetään kaikki lataukset $\lambda_{11} = \lambda_{12} = \lambda_{13} = 1$.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix},$$

missä y_1, y_2, y_3 ovat havaitut muuttujat yhdellä mittauskerralla, jossa $i = 1, 2, 3$ hypoteettinen faktori ja $j = 1, 2, 3$ mittauskerta, $\epsilon_1, \epsilon_2, \epsilon_3$ ovat kunkin indikaattorin mittausvirheet.

5.2.3 Kolmannen faktorin ensimmäisen asteen malli

Kolmannen hypoteettisen faktorin ajatellaan mittaavan ajatusten kirkastumista. Luodaan ensimmäisen kertaluvun faktorimalli niin, että jokainen aikapiste on latentti muuttuja, jolle latautuu sen aikapisteen kolmannen hypoteettisen faktorin indikaattorimuuttujat y_1, y_3, y_5 (elpyminen5) ja y_2, y_4, y_6 (elpyminen6). Skaalan asettamiseksi, kiinnitetään lataukset $\lambda_{11} = \lambda_{12} = \lambda_{13} = 1$.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{23} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix},$$

missä y_1, y_2, \dots, y_6 ovat havaitut muuttujat yhdellä mittauskerralla, λ_{ij} lataukset faktorille, jossa $i = 1, 2, 3$ hypoteettinen faktori ja $j = 1, 2, 3$ mittauskerta, $\epsilon_1, \epsilon_2, \dots, \epsilon_6$ ovat kunkin indikaattorin mittausvirheet.

5.3 Toisen kertaluvun faktorimalli hypoteettisen rakenteen pohjalta

Rakenteen tutkimiseksi ajassa, luodaan olemassa olevien ensimmäisen kertaluvun faktorimallien “yläpuolelle” kullekin yksi ylätasoinen faktori. Tarkastellaan sitten näiden ylätasoinen faktorien latauksien yhtäsuuruuksia luvussa 5.4. Ylätasoinen faktorin latauksien yhtäsuuruus viittaisi siihen, että yhden hypoteettisen faktorin rakenne ei ole muuttunut kolmen aikapisteen mittauksen aikana.

5.3.1 Mallit

Aineistosta syntyy kolme erillistä mallia, joista jokainen tarkastelee erikseen yhtä hypoteettista faktoria. Jokaisessa on siis yksi ylätasoinen faktori $F_t, t = 1, 2, 3$, jonka avulla voidaan nähdä rakenne ajassa ylätasoinen faktorien latauksista. Alemmalla tasolla on jokaisen aikapisteen latentti muuttuja $f_{ij}, i, j = 1, 2, 3$, jossa indeksi i viittaa hypoteettiseen faktoriin ja indeksi j viittaa aikapisteeseen. Latenteilla muuttujilla ovat omat indikaattorimuuttujansa, jotka ovat aineistosta saadut havaitut muuttujat hypoteettisen faktorirakenteen pohjalta.

Toisen kertaluvun mallin rakenne on kokonaisuudessaan yhtälön 3.3 mukainen. Ensimmäinen yhtälö on avattu kokonaisuudessaan luvussa 5.2, joten tarkastellaan nyt tarkemmin jälkimmäistä yhtälöä

$$(5.1) \quad \eta = \Gamma \xi + \zeta,$$

jossa η on latausmatriisi, Γ on lataus, ξ on toisen kertaluvun faktori ja ζ mittausvirhe.

Ensimmäisen kertaluvun faktorirakenteen yläpuolelle on rakennettu ylätasoinen faktori, jonka rakenne yhtälön 5.1 mukainen. Sen voi avata matriisimuodossa seuraavasti

$$\begin{bmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix},$$

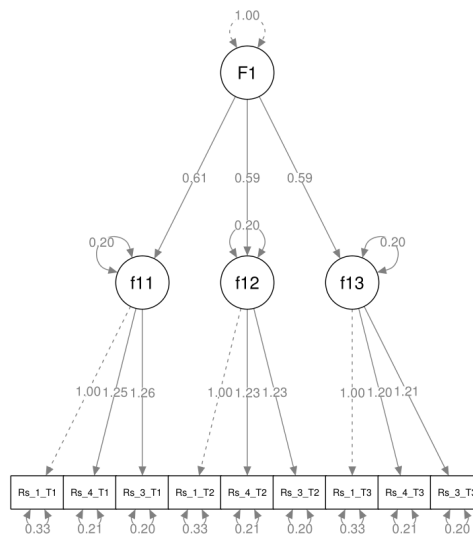
missä η_{ij} , jossa $i, j = 1, 2, 3$ on ensimmäisen kertaluvun faktori, jonka indeksit määräytyvät ensimmäisen kertaluvun mallin perusteella, γ_i , jossa $i = 1, 2, 3$ on toisen kertaluvun faktorin lataus, ξ on toisen kertaluvun faktori ja ψ_i , jossa $i = 1, 2, 3$ mittausvirhe. Kaksi keskenään verrattavaa mallia luodaan niin, että $\gamma_i, i = 1, 2, 3$ estimoituvat vapaasti tai ne asetetaan yhtä suuriksi.

Tarkastellaan rentoutumisen ja rauhallisuuden mallia. Mallin rakenteesta voidaan tehdä visuaalinen esitys. Kuvassa 5.1 F_1 on mallin ylätasoinen faktori, jonka varianssi on asetettu arvoksi yksi. F_1 :stä lähtevät nuolet ensimmäisen kertaluvun faktoreihin f_{11}, f_{12}, f_{13} , joiden varianssit on rajoitettu samaksi. Näistä kolmesta latentista muuttujasta lähtevät nuolet yhdeksään indikaattorimuuttujaan Rs_n_Tm , jossa $n = 1, 3, 4$ ja $m = 1, 2, 3$, missä n on havaitun muuttujan (elpyminen1, elpyminen3, elpyminen4) indeksi ja m aikapisteen indeksi. Indikaattorimuuttujien varianssit on rajoitettu aikapisteissä samaksi, eli yhdellä havaitulla muuttujalla on sama varianssi aikapisteiden yli. Varianssin oletetaan pysyvän samana, sillä havaitun muuttujan ajatellaan olevan

“sama muuttuja”, vaikka se olisi mitattu eri aikapisteessä. Tämä johtuu siitä, että aineiston kyselyn kysymykset sekä vastaajat pysyivät samoina kaikilla mittauskerroilla. Jokaisen ensimmäisen kertaluvun latentin muuttujan ensimmäinen indikaattorimuuttuja on kiinnitetty arvoksi yksi. Mallin muut lataukset estimoituvat vapaasti. Tarkemmat estimaatit löytyvät taulukosta 5.1.

Taulukko 5.1. Rentoutumisen ja rauhallisuuden faktorimallin lataukset ja varianssit.

Mittauskerta	Indikaattorimuuttuja	Varianssi	Lataus
f ₁₁	Restoration_1_T1	0,331	1
	Restoration_3_T1	0,203	1,264
	Restoration_4_T1	0,208	1,253
		0,203	0,613
f ₁₂	Restoration_1_T2	0,331	1
	Restoration_3_T2	0,203	1,227
	Restoration_4_T2	0,208	1,226
		0,203	0,590
f ₁₃	Restoration_1_T3	0,331	1
	Restoration_3_T3	0,203	1,206
	Restoration_4_T3	0,208	1,203
		0,203	0,587



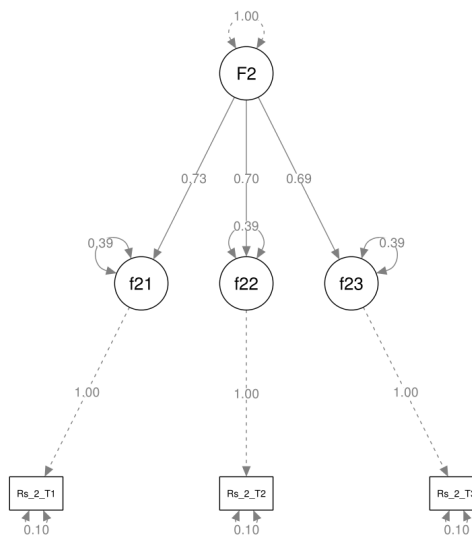
Kuva 5.1. Graafinen esitys rentoutumista ja rauhallisuutta mittaavasta faktorista kolmessa aikapisteessä.

Taulukosta 5.1 nähdään, että f₁₁, f₁₂ ja f₁₃ varianssin estimaatti on 0,203. Havaittujen muuttujien elpyminen1, elpyminen3 ja elpyminen4 varianssit ovat 0,331, 0,203 sekä 0,208. Skaalan asettaa elpyminen1, jonka lataus on kiinnitetty arvoksi

yksi. Muiden indikaattorimuuttujien elpyminen³ ja elpyminen⁴ lataukset ovat kussakin aikapisteessä hyvin lähellä toisiaan – ensimmäisellä mittauskerralla 1,264 sekä 1,253, toisella 1,227 sekä 1,226 ja kolmannella 1,206 sekä 1,203. Tämä tarkoittaa, että kun latentti muuttuja muuttuu yhden yksikön verran, havaitun muuttujan odotetaan muuttuvan latauksen verran. Sama pätee toisen kertaluvun faktorille, eli kun F_1 muuttuu yhden yksikön verran, odotetaan f_{11} , f_{12} ja f_{13} muuttuvan vastaavasti 0,613, 0,590 ja 0,587 yksikön verran.

Keskittymiskyvyn elpymistä mittaavan faktorin pohjalta tehty malli sisältää ainoastaan yhden indikaattorimuuttujan jokaista ensimmäisen kertaluvun faktorista kohti. Tämän vuoksi identifioitavuus on suurempi ongelma, kuin muissa tutkimuksen malleissa. Estimoitaville parametreille ei ole mahdollista löytää yksiselitteisiä ratkaisuja ilman tiukempia rajoituksia. Mallin graafinen esitys on esitetty kuvassa 5.2.

Tälle mallille ovat voimassa samat rajoitteet kuin rentoutumisen ja rauhallisuuden mallille, eli ylätason faktorin F_2 varianssi on kiinnitetty arvoksi yksi ja ensimmäisen kertaluvun faktorien f_{21} , f_{22} , f_{23} varianssit on asetettu samaksi. Aiemmasta poiketen indikaattorimuuttujien varianssi on asetettu arvoksi 0,1. Tämä vapauttaa estimoinnista parametreja, jolloin on mahdollista estimoida toisen kertaluvun faktorin lataukset rakenteen tutkimiseksi. Mallin lataukset ja varianssit näkyvät tarkemmin taulukosta 5.2.



Kuva 5.2. Polkukaavio keskittymiskyvyn elpymistä mittaavasta faktorista kolmessa aikapisteessä.

Taulukko 5.2. Keskitymiskyvyn elpymisen faktorimallin lataukset ja varianssit.

Mittauskerta	Indikaattorimuuttuja	Varianssi	Lataus
f ₂₁		0,1	0,729
	Restoration_2_T1	0,395	1
f ₂₂		0,1	0,700
	Restoration_2_T2	0,395	1
f ₂₃		0,1	0,686
	Restoration_2_T3	0,395	1

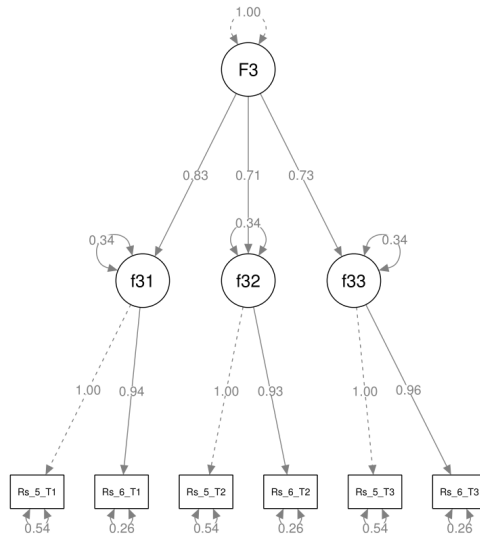
Ajatusten kirkastumista mittaavan faktorin pohjalta luotu malli on hyvin samankaltainen kuin ensimmäinen, rentoutumisen ja rauhallisuuden pohjalta luotu malli. Ylätason faktorin F₃ varianssi on kiinnitetty arvoksi yksi, alemman tason faktorien f₃₁, f₃₂ ja f₃₃ varianssi on kiinnitetty samaksi. Kullakin alemman tason faktorilla on yksi lataus kiinnitetty arvoksi yksi, jotta skaalasta tulee yksiselitteinen. Mallin muut lataukset estimoituvat vapaasti. Indikaattorimuuttujien varianssi on kiinnitetty samaksi ajan suhteen, eli samalla indikaattorilla on sama varianssi eri aikapisteissä. Parametrien tarkemmat arvot löytyvät taulukosta 5.3 ja mallin graafinen esitys kuvasta 5.3.

Taulukko 5.3. Ajatusten kirkastumisen faktorimallin lataukset ja varianssit.

Mittauskerta	Indikaattorimuuttuja	Varianssi	Lataus
f ₃₁		0,339	0,832
	Restoration_5_T1	0,536	1
	Restoration_6_T1	0,264	0,943
f ₃₂		0,339	0,715
	Restoration_5_T2	0,536	1
	Restoration_6_T2	0,264	0,926
f ₃₃		0,339	0,728
	Restoration_5_T3	0,536	1
	Restoration_6_T3	0,264	0,958

5.4 Rakenteen testaaminen

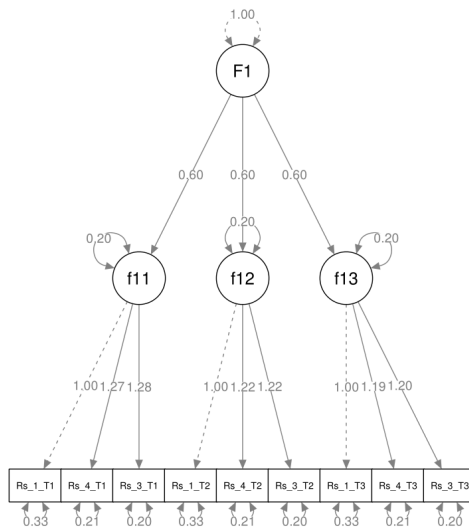
Tässä tutkielmassa faktorirakenteen pysyvyyden testaamiseksi on valittu uskottavuussuhdetesti. Aikaisemmin on määritetty kolme faktorimallia, kukin yhdelle ROS-mittarin faktorille. Näissä malleissa tärkeimmät estimoitavat parametrit ovat ylätason faktorin lataukset. Näitä malleja verrataan malleihin, joissa nämä lataukset on oletettu yhtä suuriksi. Tällöin ajatellaan, että faktorirakenne on ajan suhteen sama. Uskottavuussuhdetestillä avulla mallien sopivuutta aineistoon voidaan verrata ja nähdä,



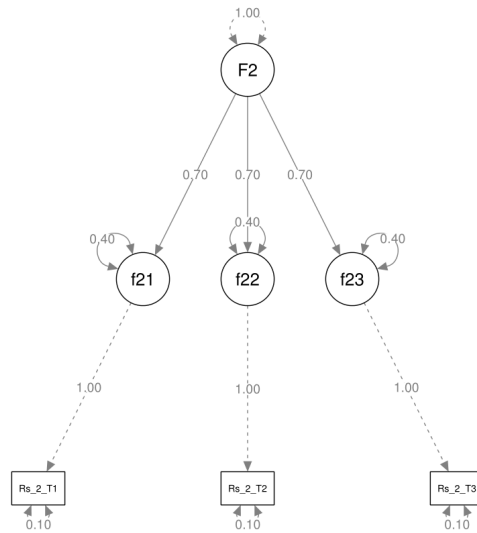
Kuva 5.3. Polkukaavio ajatusten kirkastumista mittaavasta faktorista kolmessa aikapisteessä.

onko yhtäsuuruusoletus varteenotettava vaihtoehto kyseisen teoreettisen faktorin tilanteessa.

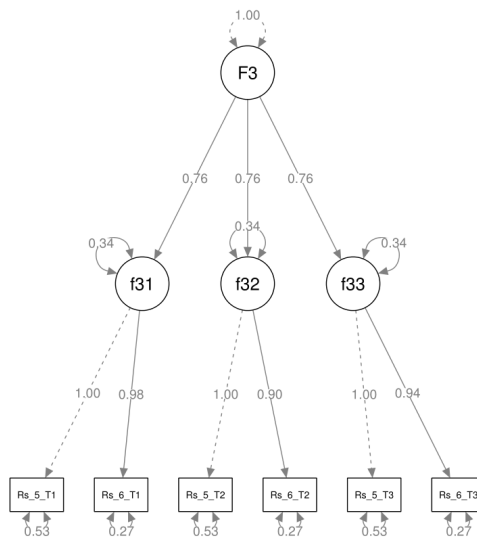
Kuvissa 5.4 – 5.6 ovat yhtäsuuruusoletuksella tehtyjen faktorimallien graafiset esitykset. Näissä kuhunkin ylätasoon faktorisiin (F_1 , F_2 , F_3) menevä lataus on kiinnitetty yhtä suureksi.



Kuva 5.4. Polkukaavio rentoutumista ja rauhallisuutta mittaavan faktorin mallista kolmessa aikapisteessä, jossa oletetaan toisen kertaluvun faktorin latauksien yhtäsuuruus.



Kuva 5.5. Polkukaavio keskittymiskyvyn elpymistä mittaavan faktorin mallista kolmessa aikapisteessä, jossa oletetaan toisen kertaluvun faktorin latauksien yhtäsuuruus.



Kuva 5.6. Polkukaavio ajatusten kirkastumista mittaavan faktorin mallista kolmessa aikapisteessä, jossa oletetaan toisen kertaluvun faktorin latauksien yhtäsuuruus.

Yhtäsuuruusoletukset sisältävien mallien tarkemmat lataukset löytyvät taulukoista 5.4 – 5.6. Rajoitettujen mallien lataukset ovat samaa suuruusluokkaa vapaasti esitöimöituvien latauksien kanssa.

Taulukko 5.4. Ensimmäisen teoreettisen faktorin yhtäsuuruusoletuksen sisältävän mallin faktorien lataukset sekä varianssit.

Mittauskerta	Indikaattorimuuttuja	Varianssi	Lataus
f ₁₁		0,203	0,597
	Restoration_1_T1	0,331	1
	Restoration_3_T1	0,203	1,278
	Restoration_4_T1	0,207	1,267
f ₁₂		0,203	0,597
	Restoration_1_T2	0,331	1
	Restoration_3_T2	0,203	1,222
	Restoration_4_T2	0,207	1,221
f ₁₃		0,203	0,597
	Restoration_1_T3	0,331	1
	Restoration_3_T3	0,203	1,197
	Restoration_4_T3	0,207	1,195

Taulukko 5.5. Toisen teoreettisen faktorin yhtäsuuruusoletuksen sisältävän mallin faktorien lataukset sekä varianssit.

Mittauskerta	Indikaattorimuuttuja	Varianssi	Lataus
f ₂₁		0,395	0,705
	Restoration_2_T1	0,1	1
f ₂₂		0,395	0,705
	Restoration_2_T2	0,1	1
f ₂₃		0,395	0,705
	Restoration_2_T3	0,1	1

Taulukko 5.6. Kolmannen teoreettisen faktorin yhtäsuuruusoletuksen sisältävän mallin faktorien lataukset sekä varianssit.

Mittauskerta	Indikaattorimuuttuja	Varianssi	Lataus
f ₃₁		0,342	0,759
	Restoration_5_T1	0,535	1
	Restoration_6_T1	0,265	0,979
f ₃₂		0,342	0,759
	Restoration_5_T2	0,535	1
	Restoration_6_T2	0,265	0,900
f ₃₃		0,342	0,759
	Restoration_5_T3	0,535	1
	Restoration_6_T3	0,265	0,938

Käytetään uskottavuussuhdetestaamiseen R-ohjelmistoa ja pakettia *lrtest*. Tästä paketista löytyy funktio *lrtest()*, johon asetetaan parametreiksi kaksi mallia, joita verrataan keskenään. Mallien täytyy olla sisäkkäiset testin onnistumiseksi. Asetetaan ensimmäiseksi parametriksi malli, jossa oletetaan ylätasoinen faktorin latauksien yhtäsuuruus, ja toiseksi parametriksi malli, jossa lataukset estimoituvat vapaasti.

Tehdään uskottavuussuhdetestaus kullekin kolmelle parille faktorimalleja. Testauksen hypoteesit ovat kuten luvussa 4, eli

$$H_0 : \Sigma = \Lambda_* \Omega_* \Lambda_*^T + \Theta_*,$$

$$H_1 : \Sigma = \Lambda \Omega \Lambda^T + \Theta,$$

missä alaindeksi * viittaa malliin, jossa lataukset on asetettu yhtäsuuriksi. Testauksen tulokset ovat seuraavanlaiset:

Taulukko 5.7. Rentoutumista ja rauhallisuutta mittaavan faktorin uskottavuussuhdetestin tulokset.

Malli	Vapausasteet	logLik	χ^2 -testisuure	p-arvo
Yhtäsuuruus	11	-6126,7		
Vapaa estimointi	13	-6126,6	0,3705	0,8309

Taulukko 5.8. Keskittymiskyvyn elpymistä mittaavan faktorin uskottavuussuhdetestin tulokset.

Malli	Vapausasteet	logLik	χ^2 -testisuure	p-arvo
Yhtäsuuruus	2	-2521,4		
Vapaa estimointi	4	-2521,0	0,8854	0,6423

Taulukko 5.9. Ajatusten kirkastumista mittaavan faktorin uskottavuussuhdetestin tulokset.

Malli	Vapausasteet	logLik	χ^2 -testisuure	p-arvo
Yhtäsuuruus	7	-5071,2		
Vapaa estimointi	9	-5068,9	4,5358	0,1035

Taulukossa 5.7 on rentoutumista ja rauhallisuutta mittaavan faktorin mallien uskottavuussuhdetestin tulokset, taulukossa 5.8 keskittymiskyvyn lepymistä mittaavan faktorin mallien testitulokset ja taulukossa 5.9 ajatusten kirkastumista mittaavan faktorin testitulokset. Taulukoista näkee mallien vapausasteet, uskottavuussuhdetestisuureen (logLik), χ^2 -testisuureen sekä tähän viittaavan *p*-arvon. Nollahypoteesi oli, että aineistoon paremmin sopiva malli on se, jossa ylätasoinen faktorien yhtäsuuruusoletus on voimassa. Tämä viittaisi siihen, että tämän testin nojalla faktorirakenne pysyy samana ajassa, kun faktoreita tarkastellaan ajassa erikseen. Nollahypoteesin

testaaminen tapahtuu χ^2 -testisuureen ja sitä vastaavan p -arvon avulla. Jos p -arvo on suurempi kuin haluttu merkitsevyysaste, niin nollahypoteesi jää voimaan.

Testeistä saadaan ensimmäisen, toisen ja kolmannen teoreettisen faktorirakenteen testaukselle χ^2 -testiarvoiksi 0,3705, 0,8854 ja 4,5358, ja näitä vastaaviksi p -arvoiksi 0,8309, 0,6423 ja 0,1035. Tällöin nollahypoteesi jää voimaan kaikilla perinteisillä merkitsevyystasoilla niin, että kolmannen eli ajatusten kirkastumista mittaavan faktorin p -arvo olisi lähimpänä nollahypoteesin hylkäämistä 10 % merkitsevyystasolla. Uskottavuussuhdetestin perusteella voidaan todeta, että yksittäin ajassa tarkasteltuna faktorirakenne on pysyvä jokaisen kolmen ROS-mittarin faktorin kohdalla.

6 Yhteenveto

Korpela ym. (2008) loivat aikaisempaan tutkimukseen pohjautuen elpymistä kuvaavan ROS-mittarin, jota mitataan kuudella kysymyksellä, joihin vastataan 7-portaisella Likert-asteikolla. Tutkielman päätavoite oli tutkia, pysyykö ROS-mittarin faktorirakenne samana yli ajan. Aineistona toimii työntekijäkyselyn vastaukset, minkä tavoitteena oli tutkia luonnolle altistumisen vaikutusta työntekijöiden hyvinvointiin. Kysely toteutettiin kolme kertaa vuoden välein.

Faktorirakenne muodostettiin Korpelan ym. (2008) tutkimuksen pohjalta. ROS-mittarin ajateltiin muodostuvan kolmesta osasta ei-havaitusta osasta: rentoutuminen ja rauhallisuus, keskittymiskyvyn elpyminen sekä ajatusten kirkastuminen. Näistä muodostettiin yksitellen faktorimalli ajan suhteen niin, että latentteina muuttujina on ei-havaittu osa eri aikapisteissä, joilla indikaattorimuuttujat aikaisemman teorian pohjalta. Tälle mallille asetettiin vielä toisen kertaluvun faktori, jotta pystyttiin tarkastelemaan latentteja muuttujia ajassa. Rakenteen ajateltiin olevan pysyvä, jos nämä lataukset ovat keskenään yhtä suuria.

Faktorirakenteen testaaminen suoritettiin suurimman uskottavuuden menetelmää hyödyntäen. Jokaiselle ei-havaitulle konseptille luotiin kaksi faktorimallia edellä mainituin tavoin - yksi niin, että toisen kertaluvun lataukset estimoituivat vapaasti, ja toinen niin, että lataukset oletetaan yhtä suuriksi. Näiden kahden mallin sopivuutta aineistoon testattiin uskottavuussuhdetestillä. Testi perustuu normaolioletukseen, jota tässä tutkielmassa käytetty Likert-asteikollinen aineisto ei täytä. Tämä voi vaikuttaa tulosten luotettavuuteen.

Uskottavuussuhdetestin tuloksena oli, että yksitellen ajassa tarkasteltuna, alkuperäisen teorian latenttien muuttujien faktorirakenne on pysyvä yli ajan. Jatkotutkimuskohteena voisi olla yhdistää nämä kolme erillistä mallia, ja testata rakenteen pysyvyyttä samaan tapaan, ja sallien esimerkiksi latenttien muuttujien välisiä kovariansseja.

Lähteet

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Vol. 210. Wiley Series in Probability and Statistics Ser. John Wiley and Sons, Inc. EBook Central: 9781118619162.
- Jöreskog, K. G. ja D. Sörbom (1979). *Advances in Factor Analysis and Structural Equation Modeling*. Abt. Associates Inc., s. 159–161. ISBN: 0-89011-535-4.
- Korpela, K. M., T. Nummi et al. (2017). ”Nature exposure predicts employee well-being trajectory groups across two years”. Englanti. *Journal of Environmental Psychology*.
- Korpela, K. M., M. Ylén et al. (2008). ”Determinants of restorative experiences in everyday favourite places”. Englanti. *Health & Place* 14.
- Leskinen, E. (1987). *FAKTORIANALYYSI. Konfirmatoristen faktorimallien teoria ja rakentaminen*. Jyväskylän yliopiston monistuskeskus.