Structural Equation Modelling

Dr. Calum Webb

Sheffield Methods Institute, the University of Sheffield.

c.j.webb@sheffield.ac.uk

Pre-requisites

If I want to learn Structural Equation Modelling I should already have...

- A very good understanding of multiple linear regression
- Good familiarity with terms like variance, covariance, correlation
- It can help to have some familiarity with data tidying (i.e. in R), especially for converting between long and wide data formats
- Formal understanding of causal diagrams/Directed Acyclic Graphs can be helpful

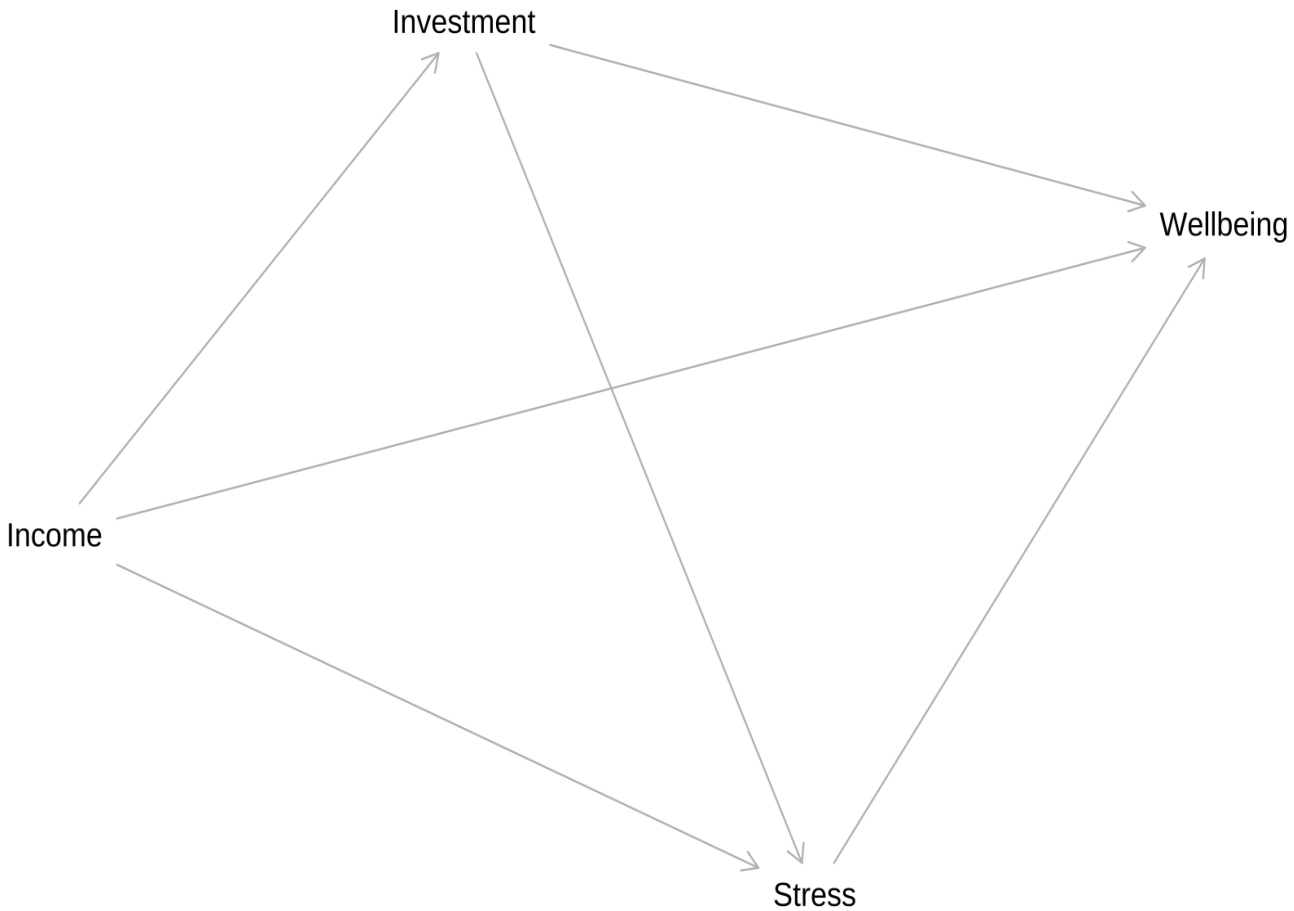
Is it more effective to increase family income and spending on their children to improve child wellbeing, or is it more effective to reduce family stress?

Variables

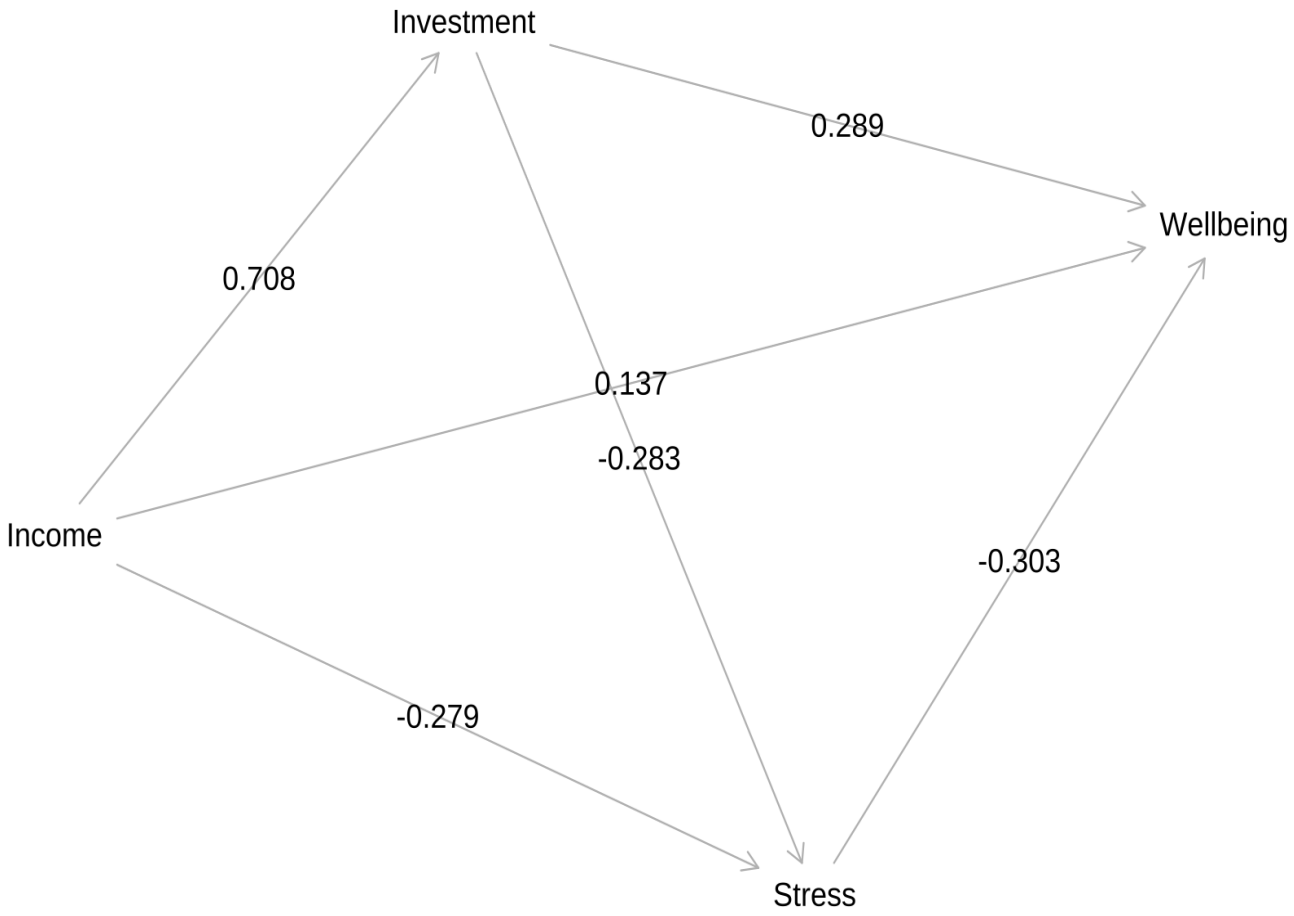
- `income` = Family income (in £1000s)
- `investment` = Amount spent on child (in £1000s)
- `stress_q1` = Approximately how many times over the last year have you gotten into a disagreement with a co-parent or family member about the parenting of your child(ren)? (10 point scale, centered at 5)
- `scws` = Child wellbeing (Stirling Child Wellbeing Scale, range 12-60, higher = better wellbeing)

Is it more effective to increase family income and spending on their children to improve child wellbeing, or is it more effective to reduce family stress?

	Dependent variable:
	scale(scws)
scale(income)	0.137*** (0.016)
scale(investment)	0.289*** (0.016)
scale(stress_q1)	-0.303*** (0.013)
Constant	0.000 (0.011)
Observations	5,000
R^2	0.374
Adjusted R^2	0.374
Residual Std. Error	0.791 (df = 4996)
F Statistic	994.657*** (df = 3; 4996)
Note:	*p<0.1; **p<0.05; ***p<0.01

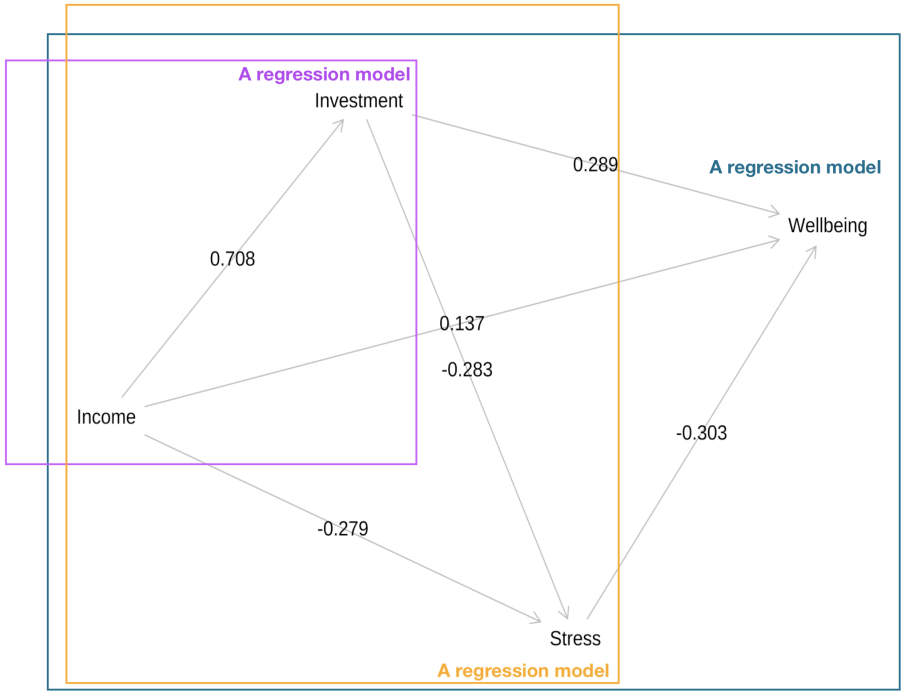


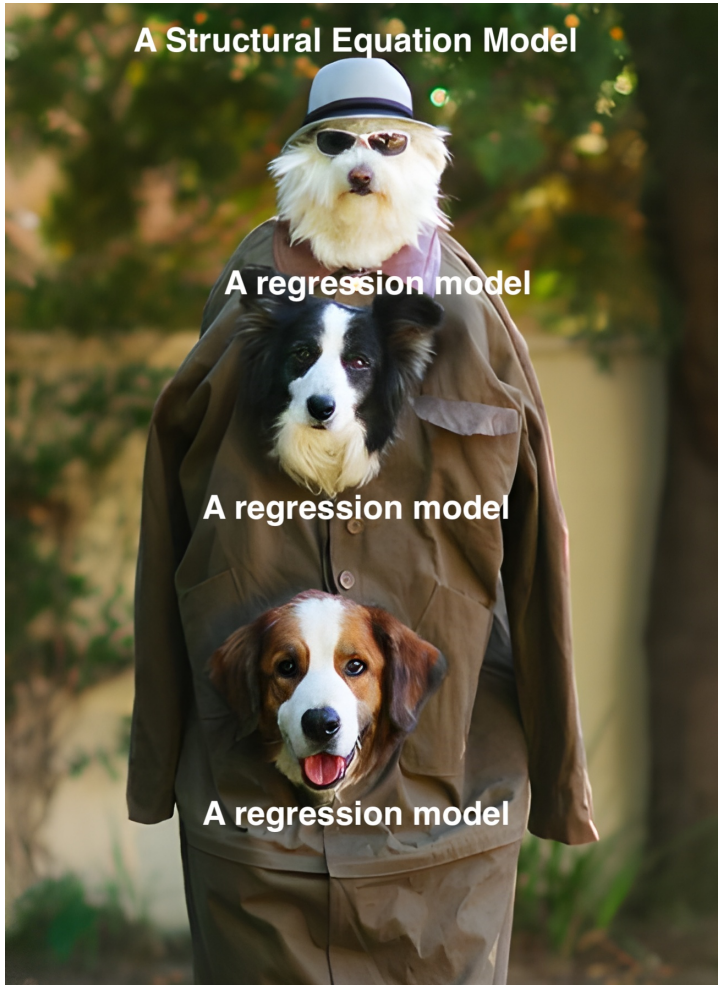
- Can't have investment without income...
- Income and investment probably both impact family stress, which then impacts wellbeing...


Investment
0.289
Wellbeing
0.708
0.137
-0.283
Income
-0.303
-0.279
Stress

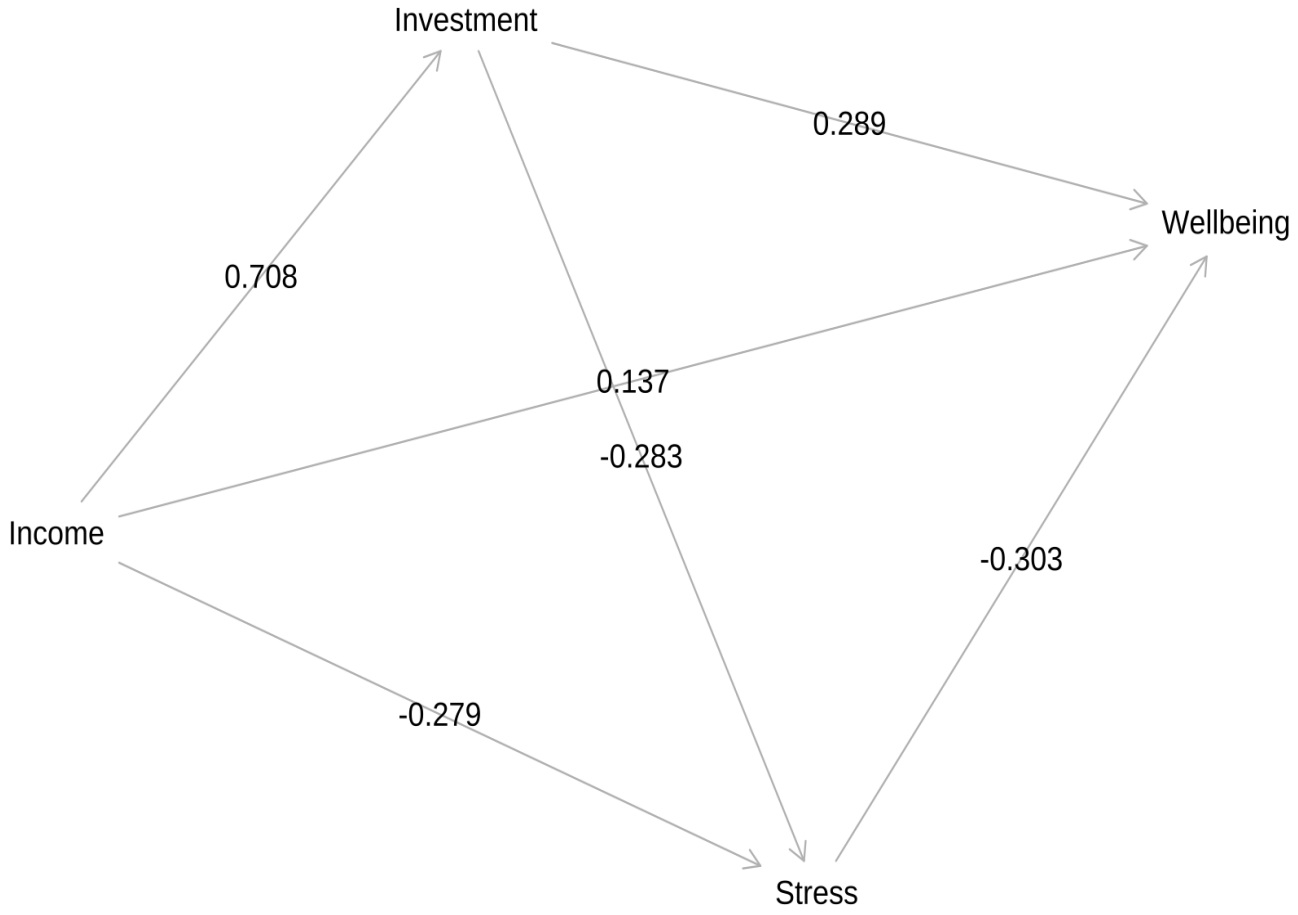
Structural Equation Modelling

```
Wellbeing ~ Income + Investment + Stress

Stress ~ Income + Investment

Investment ~ Income
```





Path Analysis



Direct Effects

- Income: 0.137
- Investment: 0.289
- Stress: -0.303

Indirect Effects

Income through...

- Stress: -0.279 x -0.303 = 0.085
- Investment: 0.708 x 0.289 = 0.204
- Investment & Stress: 0.708 x -0.283 x -0.303 = 0.061

Total Income Effect

0.137 + 0.085 + 0.204 + 0.061 = 0.487

But what is family stress? And is this question the best measure of it?

Latent Variables (Factor Analysis)

- We can't just ask: "How much family stress do you experience on a scale of 1-10?" and expect a good measure.
- We operationalise the concept into distinct questions, e.g.:
    - How frequently do you have disagreements about parenting?
    - Do you sometimes regret having children?
    - How often do you feel like you go to bed angry at your partner?
- All of these questions capture *something* about family stress, but none of them capture it exactly.

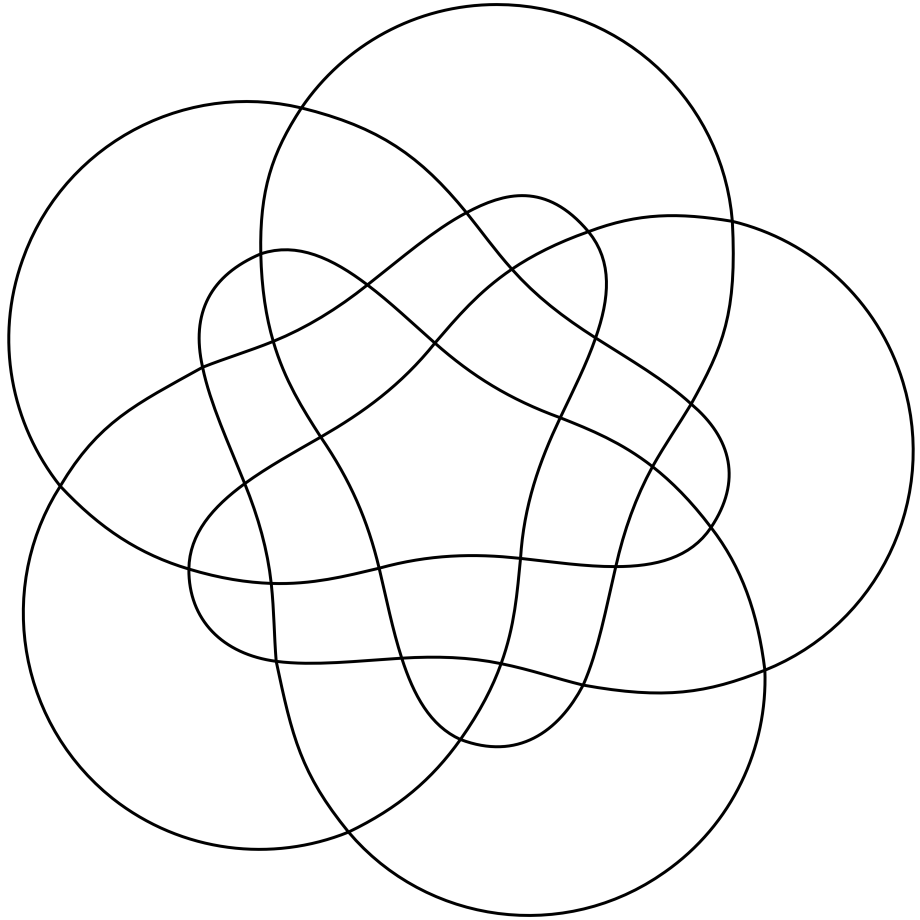
Latent Variables (Factor Analysis)

We shouldn't

- Chuck all of these similar questions into a single regression model (because of multicollinearity)
- Sum up all of the responses uncritically:
    - What if they are all on different scales and need to be weighted differently?
    - What if some of the questions are worse measures of the concept than others?
    - What if some questions are more salient for some groups of people than others?

We can

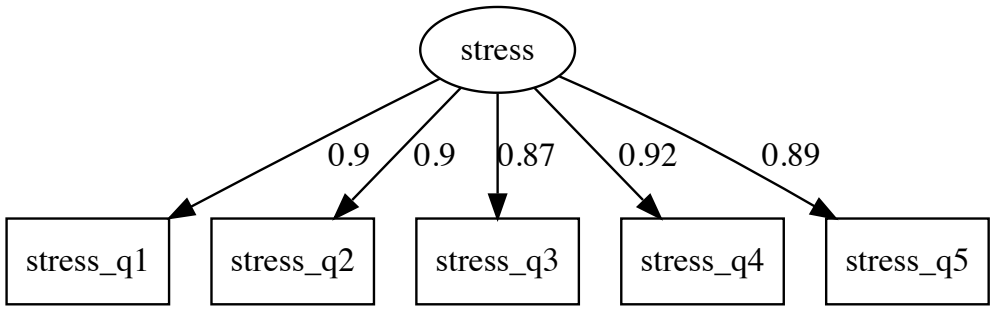
- Use factor analysis/construct a latent variable to try and capture the 'underlying' concept.



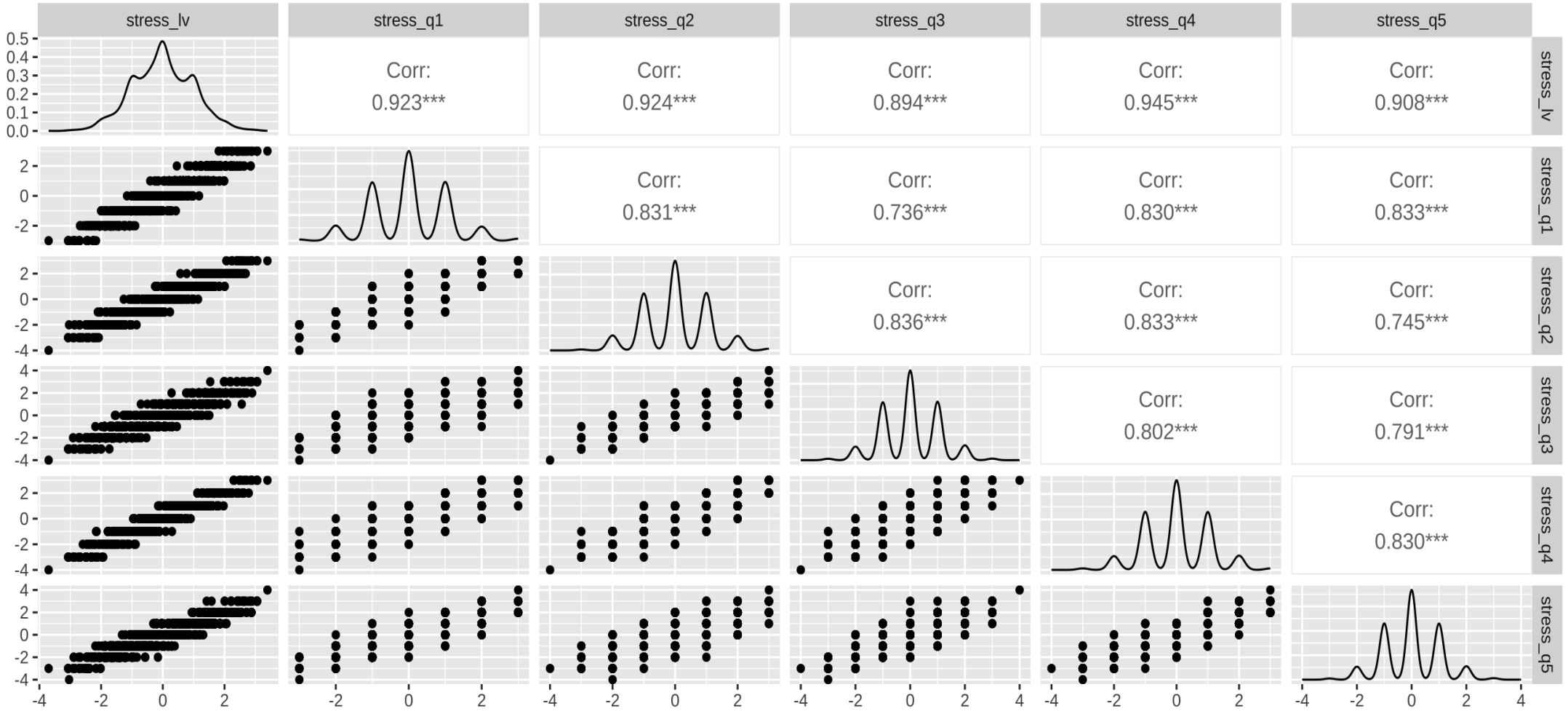
Latent Variables (Factor Analysis)

Factor analysis constructs a latent variable that simultaneously predicts multiple indicator (or manifest) variables.

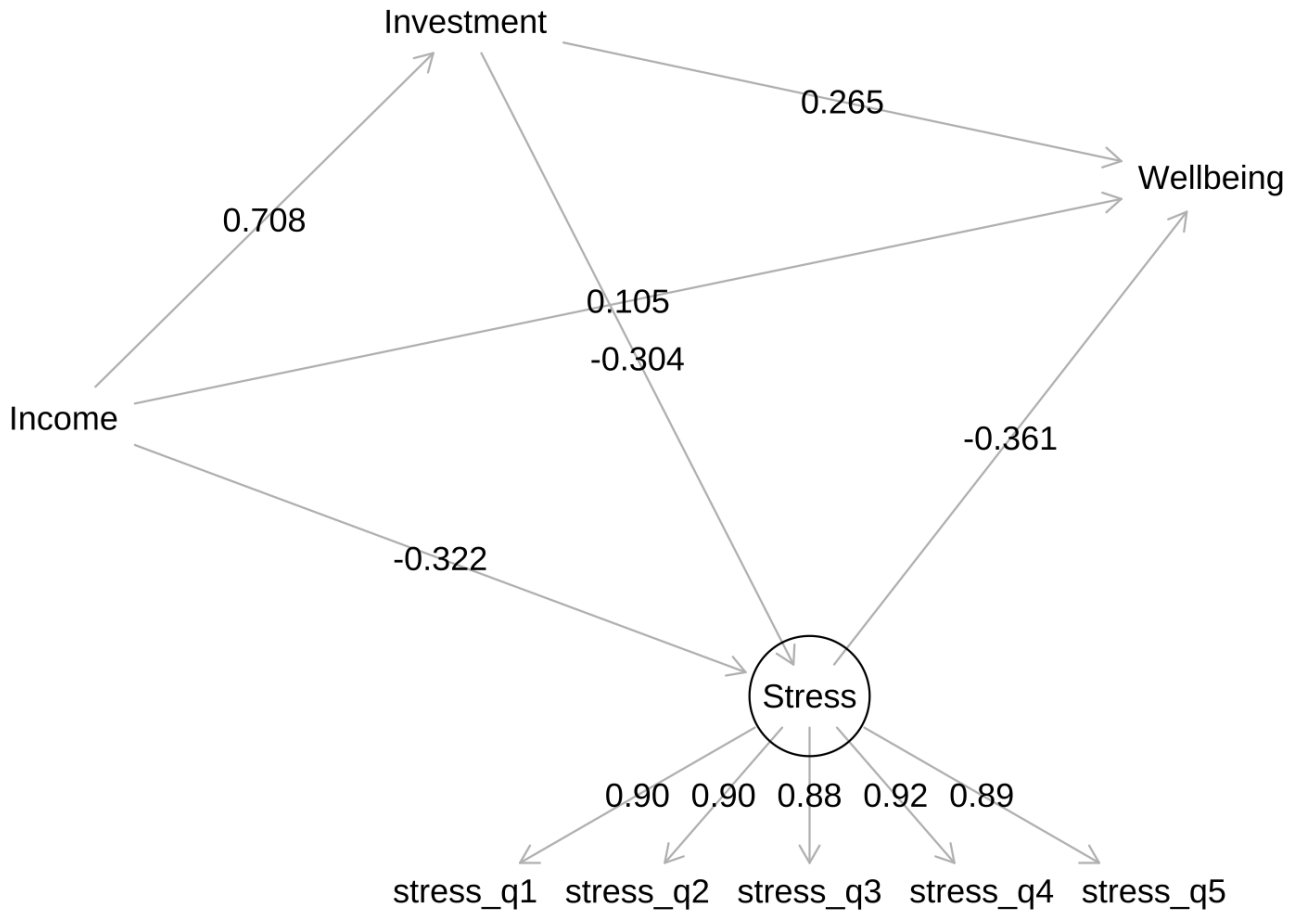
Its scale is arbitrary but is commonly fixed to either a) a marker variable or b) a standardised distribution (mean = 0, sd = 1).



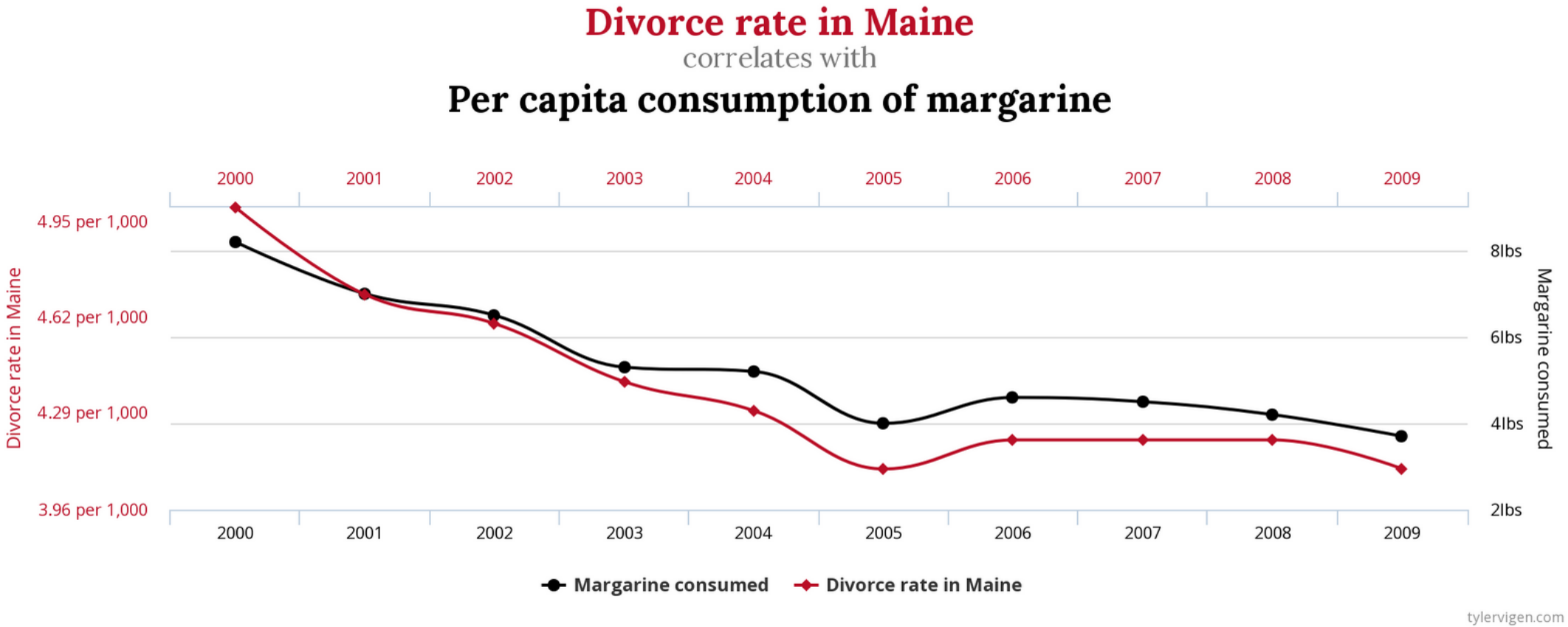
Latent Variables (Factor Analysis)



SEM with Latent Variables

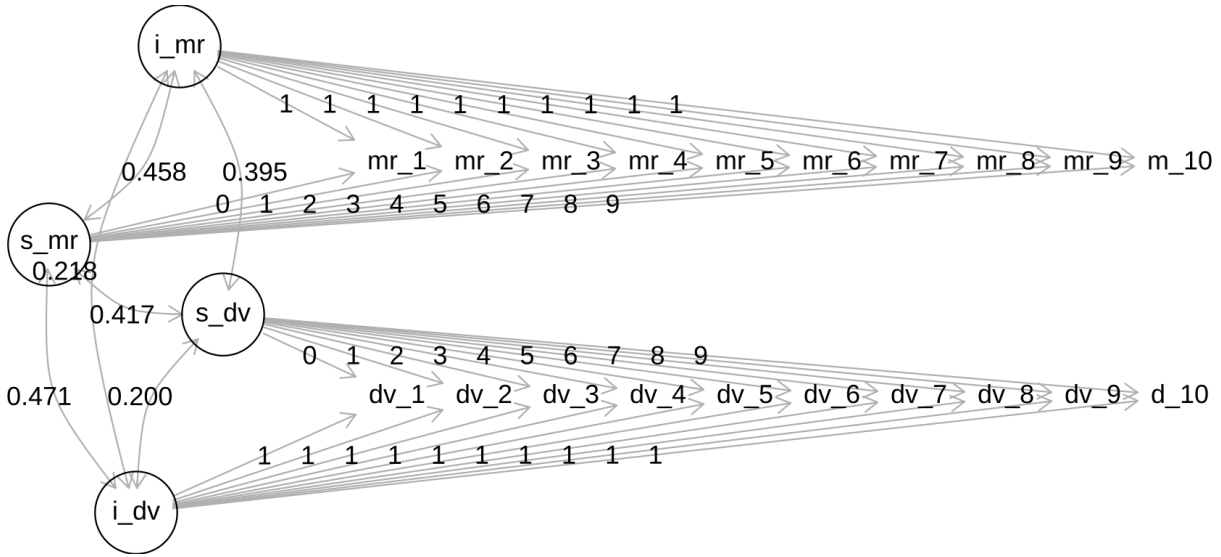


Latent Growth Modelling



Latent Growth Modelling

Constraints on latent variables can also be used in clever ways to capture change over time.

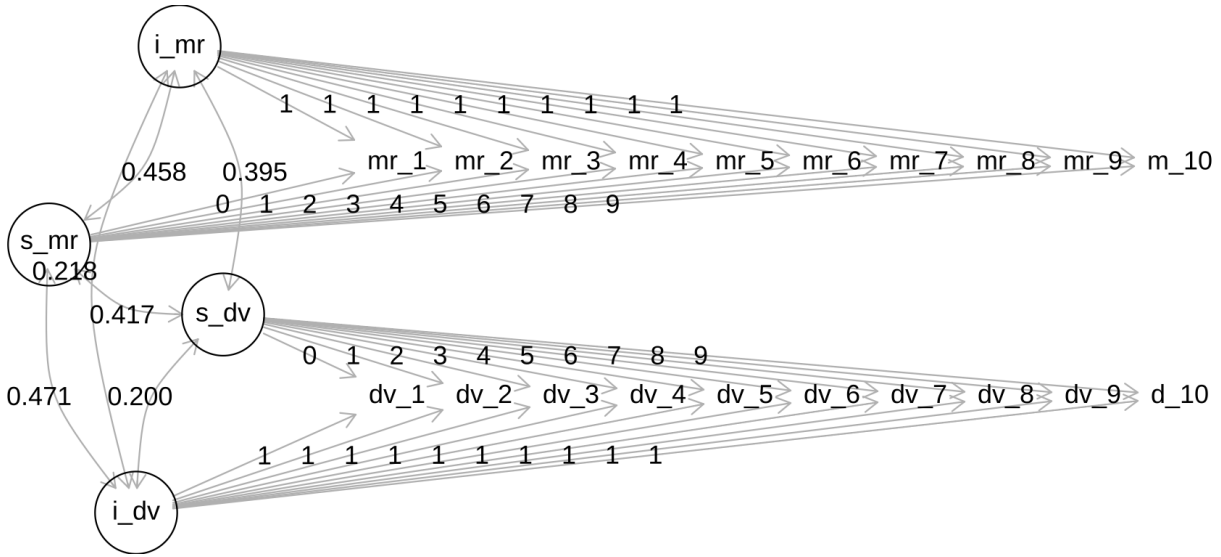


Latent Growth Modelling

Constraints on latent variables can also be used in clever ways to capture change over time.

- The mean linear slope for divorce rates was -0.203 per year
- The mean linear slope for margarine consumption was -0.215 per year, explaining much of why the two are correlated.

But, a positive correlation between state-level slopes in divorce rate and state-level slopes in margarine consumption (0.417) suggests that divorce rates were falling faster in states where margarine consumption was falling faster (and vice-versa)



Latent Growth Modelling

The fact that latent growth models use wide rather than long data means that there are a few more nuanced options for examining associations between trends as well as residuals between trends:

- What are the associations between latent traits/individuals: intercepts
- What are the associations between latent trends: slopes
- What are the associations between time varying residuals: indicator variable residuals

Open access taster workbook on SEM, CFA, and LGM in `R` using the `lavaan` package

Link: https://github.com/cjrwebb/aqm-taster-sem

- Click on Code -> Download as .zip -> Open .Rproj file and `worked-examples.R`