Factor Models: Specification, Identification and Estimation

Summer school in Bayesian econometrics – Part 10

Rémi Piatek

Department of Economics
University of Copenhagen

General Introduction

Spearman (1904): Model of **g**eneral intelligence theory (‘g’)

- Tests scores of 36 boys on following topics: classics, French, English, mathematics, discrimination of tones, musical talent.
- Found a common factor underlying these test scores, the ‘g’ factor:

$$y_j = a_j g + \psi_j$$

See Stanley A. Mulaik (2009), Foundations of Factor Analysis, 2nd ed., Chapman & Hall/CRC.

In many scientific fields, researchers have to deal with unobserved constructs.

Some examples:

- **Biology:** cell resistance to a chemical compound
- **Genetics:** specific trait related to a set of genes
- **Psychology:** cognitive abilities, personality traits
- **Economics:** time preferences, risk aversion
- **Marketing:** consumer preferences and tastes

In many scientific fields, researchers have to deal with unobserved constructs.

Some examples:

- **Biology:** cell resistance to a chemical compound
- **Genetics:** specific trait related to a set of genes
- **Psychology:** cognitive abilities, personality traits
- **Economics:** time preferences, risk aversion
- **Marketing:** consumer preferences and tastes

Factor analysis as a data reduction method:
Makes it possible to extract a small number of latent factors from a larger set of response variables.

Some examples

Factor analysis has become very popular in the social sciences, and especially in psychology:

- 'g' factor
- Big Five (OCEAN): Openness to new experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism.

More recently, economists have started using it to incorporate psychological concepts in their theoretical frameworks:

- Cognitive abilities
- Personality traits
- Time preferences
- Risk aversion
- ...

See Heckman and coauthors for examples.

Let us imagine the following problem...

You want to construct an economic model and control for some unobserved personal characteristics like cognitive abilities and self-esteem.

For this purpose, you extract two latent factors from a set of 15 items related to these latent traits.

Cognitive tests — The ASVAB consists of a battery of tests that measure knowledge and skills in the following areas:

1. Arithmetic reasoning
2. Word knowledge
3. Paragraph comprehension
4. Numerical operations
5. Mathematics knowledge

Rosenberg's self-esteem scale (1965):

1. I feel that I'm a person of worth, at least on an equal basis with others.
2. I feel that I have a number of good qualities.
3. All in all, I am inclined to feel that I am a failure.
4. I am able to do things as well as most other people.
5. I feel I do not have much to be proud of.
6. I take a positive attitude toward myself.
7. On the whole, I am satisfied with myself.
8. I wish I could have more respect for myself.
9. I certainly feel useless at times.
10. At times I think I am no good at all.

Table 1: Results from factor analysis

	θ_1	θ_2	Uniq.
ASVAB 1	0.926	-0.063	0.188
ASVAB 2	0.850	0.045	0.243
ASVAB 3	0.843	0.036	0.263
ASVAB 4	0.753	0.030	0.414
ASVAB 5	0.887	-0.056	0.250
Rosenberg 1	0.065	0.583	0.624
Rosenberg 2	-0.013	0.656	0.577
Rosenberg 3	0.100	0.630	0.541
Rosenberg 4	-0.054	0.626	0.633
Rosenberg 5	0.106	0.580	0.602
Rosenberg 6	-0.087	0.741	0.497
Rosenberg 7	-0.111	0.605	0.677
Rosenberg 8	0.147	0.366	0.799
Rosenberg 9	-0.025	0.499	0.760
Rosenberg 10	-0.067	0.582	0.689

How were these results produced? How can we interpret them?

Model specification

Factor model with p observed response variables and k latent factors ($k \ll p$):

$$\begin{aligned}
y_{1i} &= \alpha_{11}\theta_{1i} + \alpha_{12}\theta_{2i} + \ldots + \alpha_{1k}\theta_{ki} + \varepsilon_{1i} \\
y_{2i} &= \alpha_{21}\theta_{1i} + \alpha_{22}\theta_{2i} + \ldots + \alpha_{2k}\theta_{ki} + \varepsilon_{2i} \\
&\vdots \\
y_{pi} &= \alpha_{11}\theta_{1i} + \alpha_{p2}\theta_{2i} + \ldots + \alpha_{pk}\theta_{ki} + \varepsilon_{pi}
\end{aligned}$$

for each individual $i = 1, ..., n$.

Model in compact form:

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{pi} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pk} \end{pmatrix} \begin{pmatrix} \theta_{1i} \\ \theta_{2i} \\ \vdots \\ \theta_{ki} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{pi} \end{pmatrix}$$

or simply:

$$y_i = \alpha\theta_i + \varepsilon_i$$

α is the matrix of factor loadings, ε is a vector of error terms.

Latent factors:

$$\theta_i \sim N(0, \boldsymbol{\Psi}_\theta)$$

where:

$$\boldsymbol{\Psi}_\theta = \begin{pmatrix} \psi_1^2 & \psi_{12} & \dots & \psi_{1k} \\ \psi_{21} & \psi_2^2 & \dots & \psi_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{k1} & \psi_{k2} & \dots & \psi_k^2 \end{pmatrix} \qquad \text{or} \qquad \boldsymbol{\Psi}_\theta = \mathrm{diag}(\psi_1^2, \dots, \psi_k^2)$$

(correlated case) (uncorrelated case)

Error terms:

$$\varepsilon_i \sim N(0, \boldsymbol{\Sigma}_\varepsilon) \qquad \boldsymbol{\Sigma}_\varepsilon = \mathrm{diag}(\sigma_1^2, \dots, \sigma_p^2)$$

The **error terms** are assumed to be **mutually independent** and **independent of the latent factors**:

$$\varepsilon_j \perp\!\!\!\perp \varepsilon_l, \forall j \neq l \qquad\qquad \varepsilon \perp\!\!\!\perp \theta$$

Conditional and unconditional distributions of the vector of response variables y_i:

$$y_i \mid \theta_i \sim N(\alpha\theta_i, \boldsymbol{\Sigma}_\varepsilon) \qquad\qquad y_i \sim N\big(0, \alpha\boldsymbol{\Psi}_\theta\alpha' + \boldsymbol{\Sigma}_\varepsilon\big)$$

When conditioning on θ_i, the covariance matrix is diagonal $\Longrightarrow$ The latent factors explain all the dependence structure among the p observed response variables.

So far we have assumed that $E(\theta) = E(\varepsilon) = 0$, implying that the response variables y are centered.

Explanatory variables x can be introduced in the model without any complications:

$$y = x\beta + \alpha\theta + \varepsilon \qquad\qquad y \mid x, \theta \sim N(x\beta + \alpha\theta, \Sigma_\varepsilon)$$

Independence assumption:

$$x \perp\!\!\!\perp \theta \perp\!\!\!\perp \varepsilon$$

Conditional on x, the covariance structure of the model is not affected by these covariates:

$$\text{cov}(y \mid x) = \alpha \Psi_\theta \alpha' + \Sigma_\varepsilon$$

The mean vector is identified by the mean of the response variables:

$$E(y) = x\beta$$

Identification

The covariance structure is **not identified**!

Different problems at stake. Identification is a **big issue** in factor modeling.

For any invertible matrix Q of dimension $(k \times k)$, the model can be redefined as:

$$\widetilde{\theta}_i = Q\theta_i \qquad \widetilde{\alpha} = \alpha Q^{-1}$$
$$V(\widetilde{\theta}_i) = Q\Psi_\theta Q'$$

This transformed model is equivalent to the initial one since:

$$V(y_i) = \widetilde{\alpha} V(\widetilde{\theta}_i)\widetilde{\alpha}' + \Sigma_\varepsilon = \alpha\Psi_\theta\alpha' + \Sigma_\varepsilon$$

Q is called **rotation matrix**.

The standard factor model assumes that $\Psi_\theta = I_k$.

This standardization rules out the previous transformation, except when Q is an orthogonal matrix of dimension $(k \times k)$:

$$\widetilde{\theta}_i = Q\theta_i \qquad \widetilde{\alpha} = \alpha Q'$$

$$V(\widetilde{\theta}_i) = Q\Psi_\theta Q'$$

Here again, the transformed model is equivalent to the initial one.

[Note: Q is an orthogonal matrix if $Q' = Q^{-1}$, which entails $Q'Q = QQ' = I_k$]

To prevent **factor rotation** in the standard factor model ($\Psi_\theta = I_k$), Geweke and Zhou (1996) suggest to use a block lower triangular factor loading matrix:

$$\alpha = \begin{pmatrix} \alpha_{11} & 0 & 0 & \dots & 0 \\ \alpha_{21} & \alpha_{22} & 0 & \dots & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{j1} & \alpha_{j2} & \alpha_{j3} & \dots & \alpha_{jk} \\ \alpha_{j+1,1} & \alpha_{j+1,2} & \alpha_{j+1,3} & \dots & \alpha_{j+1,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \alpha_{p3} & \dots & \alpha_{pk} \end{pmatrix}$$

See also Lopes and West (2004), Frühwirth-Schnatter and Lopes (2010).

Another problem: the number of parameters should not exceed the number of available equations!

Standard factor model:

$$V(y_i) = \Omega = \alpha\alpha' + \Sigma_\varepsilon$$

Covariance matrix Ω:

- Number of free parameters: $p(k+1) - k(k-1)/2$
- Number of unique elements: $p(p+1)/2$

This leads to the constraint that:

$$p(k+1) - k(k-1)/2 \leq p(p+1)/2$$
$$k \leq p + \frac{3}{2} - \sqrt{2p + \frac{9}{4}}$$

The maximum number of latent factors that can be specified is therefore limited by the number of response variables:

- $p = 6$ implies $k \leq 3$,
- $p = 7$ implies $k \leq 4$,
- $p = 10$ implies $k \leq 6$,
- $p = 20$ implies $k \leq 15$,
- $p = 100$ implies $k \leq 87$.

This problem refers to the well-known **Ledermann (1937) bound** in the literature.

Does the ordering of the observed variables in y matter?

Alternative orderings are trivially produced via:

$$\widetilde{y} = Ay$$

for some switching matrix A.

The new rotation has the same latent factors but transformed loading matrix $\widetilde{\alpha} = A\alpha$.

Problem: the transformed factor loading matrix is not necessarily block lower triangular.

We can always find an orthogonal matrix P such that

$$\alpha = \widetilde{\alpha} P' = A \alpha P'$$

is block lower triangular and common factors

$$\widetilde{\theta} = P\theta$$

still $N(0, I_k)$, see Lopes and West (2004).

$\Longrightarrow$ The order of the variables in y is immaterial when k is properly chosen, i.e. when α is full-rank.

Any latent factor and the corresponding factor loadings can be multiplied by -1 without changing the covariance structure:

$$\widetilde{\theta}_j = -\theta_j \qquad \widetilde{\alpha}_{.j} = -\alpha_{.j}$$

where $\alpha_{.j}$ is the jth column of α.

Solutions to prevent sign-switching:

- Constrain the diagonal elements of α to be positive:

$$\alpha_{jj} \geq 0, \forall j = 1, \ldots, k.$$

- Instead of normalizing the variance of the latent factors to 1, normalize the diagonal elements of α:

$$\alpha_{jj} = 1, \forall j = 1, \ldots, k.$$

Factor rotation can be used on purpose, for instance to make the interpretation of the latent factors easier.

Some examples of popular rotation methods:

- **Varimax** (Kaiser, 1958): orthogonal rotation that aims at maximizing the sum of the variances of the squared loadings, providing axes with as few large loadings and as many near-zero loadings as possible.
- **Quartimax**: orthogonal rotation minimizing the number of factors explaining each response variable.
- **Promax**: non-orthogonal rotation method.

Table 2: Standard factor analysis

	θ_1	θ_2	Uniq.
ASVAB 1	0.875	-0.217	0.188
ASVAB 2	0.862	-0.114	0.243
ASVAB 3	0.850	-0.121	0.263
ASVAB 4	0.757	-0.110	0.414
ASVAB 5	0.842	-0.204	0.250
Rosenberg 1	0.396	0.468	0.624
Rosenberg 2	0.360	0.542	0.577
Rosenberg 3	0.457	0.500	0.541
Rosenberg 4	0.303	0.525	0.633
Rosenberg 5	0.434	0.458	0.602
Rosenberg 6	0.336	0.624	0.497
Rosenberg 7	0.235	0.517	0.677
Rosenberg 8	0.354	0.275	0.799
Rosenberg 9	0.260	0.415	0.760
Rosenberg 10	0.265	0.491	0.689

Table 3: Promax rotation

	θ_1	θ_2	Uniq.
ASVAB 1	0.926	-0.063	0.188
ASVAB 2	0.850	0.045	0.243
ASVAB 3	0.843	0.036	0.263
ASVAB 4	0.753	0.030	0.414
ASVAB 5	0.887	-0.056	0.250
Rosenberg 1	0.065	0.583	0.624
Rosenberg 2	-0.013	0.656	0.577
Rosenberg 3	0.100	0.630	0.541
Rosenberg 4	-0.054	0.626	0.633
Rosenberg 5	0.106	0.580	0.602
Rosenberg 6	-0.087	0.741	0.497
Rosenberg 7	-0.111	0.605	0.677
Rosenberg 8	0.147	0.366	0.799
Rosenberg 9	-0.025	0.499	0.760
Rosenberg 10	-0.067	0.582	0.689

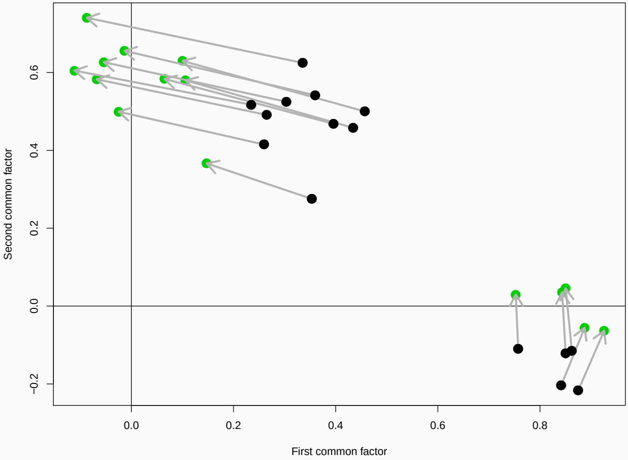


Figure 1: Promax rotation — before (black) and after (green)

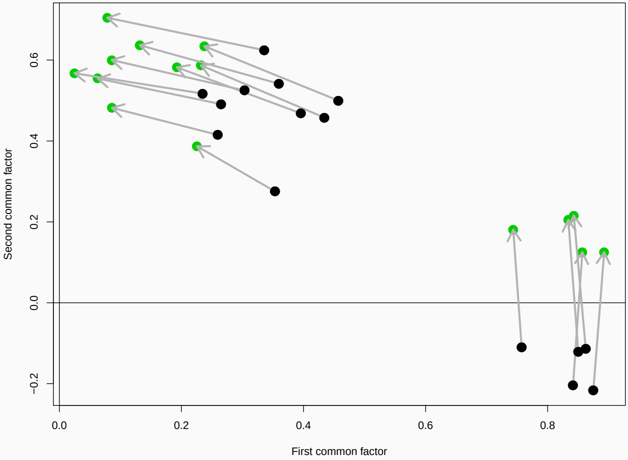


Figure 2: Varimax rotation — before (black) and after (green)

Dedicated measurements

When some prior information is available on the latent structure, **dedicated measurements** can be specified by imposing some zero restrictions on the factor loading matrix:

- Makes the latent components of the model **more interpretable**.
- **Overidentifying assumptions**, since the number of parameters to be estimated decreases while the information available from the covariance structure does not change.
- Restrictions required to **secure identification** of the model in some cases, especially when the factors are correlated.

Many empirical studies rely on dedicated measurements.

Examples: Heckman, Stixrud & Urzua (2006) specify one latent factor one cognitive abilities and one composite factor for noncognitive kills.

$$\begin{array}{ll}
\left(\begin{array}{cc}
\alpha_{11} & 0 \\
\alpha_{21} & 0 \\
\alpha_{31} & 0 \\
\alpha_{41} & 0 \\
0 & \alpha_{52} \\
0 & \alpha_{62} \\
0 & \alpha_{72} \\
\alpha_{81} & \alpha_{82} \\
\alpha_{91} & \alpha_{92}
\end{array}\right)
&
\begin{array}{l}
\\
\\
\Longleftarrow \quad \text{cognitive tests} \\
\\
\\
\Longleftarrow \quad \text{noncognitive measures} \\
\\
\\
\Longleftarrow \quad \text{outcome system}
\end{array}
\end{array}$$

Assume now that Ψ_θ is a non-diagonal covariance matrix.

- Cholesky decomposition of a symmetric, positive-(semi)definite matrix:

$$\Psi_\theta = \Gamma\Gamma'$$

  where Γ is a lower triangular matrix with strictly positive diagonal elements.

- Without further restrictions, the model with correlated factors can be re-expressed as a standard factor model since:

$$\begin{aligned}\Omega = \alpha\Psi_\theta\alpha' + \Sigma_\varepsilon &= (\alpha\Gamma)I_k(\alpha\Gamma)' + \Sigma_\varepsilon \\ &= \widetilde{\alpha}I_k\widetilde{\alpha}' + \Sigma_\varepsilon\end{aligned}$$

How to prevent factor rotation and guarantee identification?

- The block lower triangular structure of the factor loading matrix is **no longer sufficient**:

$$\Gamma \text{ lower triangular} \quad \Rightarrow \quad (\alpha\Gamma) \text{ block lower triangular}$$

- Ψ_θ is a **non-diagonal** covariance matrix

$$\Rightarrow \quad k(k-1)/2 \text{ more parameters have to be estimated.}$$

- **More restrictions** are required on the factor loading matrix

$$\Rightarrow \quad \text{Use of dedicated measurements (see Williams, 2017)}$$

Extending the standard factor model

Many **assumptions** of the standard factor model appear to be **too strong** in practice and can be relaxed:

- **Normality** assumption of the response variables (conditional on x)
- **Independence** of the latent factors
- **Normality** assumption of the latent factors
- **Normality** assumption of the error terms
- **Fixed** number of latent factors
- **Linear form** of the latent structure

Many cases in practice where the response variables are **not continuous**.

Example: Psychological tests where the answers can be of different types:

- **Binary** response: "agree" (1) or "disagree" (0).
- **Ordered** response (Likert scale): "strongly agree", "slightly agree", "slightly disagree", "strongly disagree".
- **Truncated** response: number between 0 ("disagree") and 100 ("disagree").

Misspecification of the model $\Longrightarrow$ biased estimates of model parameters and extraction of latent factors is jeopardized.

Assuming a latent underlying variable for response variable $j = 1, ..., p$, for individuals $1, ..., n$:

$$y_{ji}^{\star} = x_{ji}\beta_j + \alpha_{j}.\theta_i + \varepsilon_{ji}$$

Observed variable:

$$y_{ji} = \begin{cases} y_{ji}^{\star} & \text{continuous case} \\ \mathbb{1}[y_{ji}^{\star} > 0] & \text{dichotomous case} \\ y_{ji}^{\star}\mathbb{1}[y_{ji}^{\star} > 0] & \text{censored case} \\ c\mathbb{1}[\gamma_{c-1} \leq y_{ji}^{\star} < \gamma_c] & \text{ordinal case (for } c = 1, ..., C) \end{cases}$$

How to relax the normality assumption on the latent factors?

Kotlarski's theorem (1967)

Let W, Z_1 and Z_2 be three independent real random variables and let:

$$Y_1 = W + Z_1$$
$$Y_2 = W + Z_2$$

If the characteristic function of the pair (Y_1, Y_2) does not vanish, then the distribution of (Y_1, Y_2) determines the distributions of W, Z_1, Z_2 up to a change of the location.

$\Longrightarrow$ If the model is appropriately specified, this theorem can be sequentially applied to each pair of response variables to retrieve the marginal distributions of the respective latent factors and error terms.

Factor model:

$$
\begin{aligned}
Y_1 &= \mu_1 + \theta + \varepsilon_1 \\
Y_2 &= \mu_2 + \alpha_2\theta + \varepsilon_2 \\
Y_3 &= \mu_3 + \alpha_3\theta + \varepsilon_3
\end{aligned}
$$

Assumptions:

$$
\begin{aligned}
E(\theta) = E(\varepsilon_j) &= 0 \\
\varepsilon_j \perp\!\!\!\perp \varepsilon_{j'} \perp\!\!\!\perp \theta & \qquad\qquad \text{for } j, j' = 1, 2, 3
\end{aligned}
$$

Intercepts identified from the means of the response variables:

$$E(Y_1) = \mu_1 \qquad E(Y_2) = \mu_2 \qquad E(Y_3) = \mu_3$$

Factor loadings identified by the ratios of covariances:

$$\alpha_2 = \frac{cov(Y_2, Y_3)}{cov(Y_1, Y_3)} \qquad \alpha_3 = \frac{cov(Y_2, Y_3)}{cov(Y_1, Y_2)}$$

Given the identification of the intercepts and of the factor loadings (up to the normalization $\alpha_1 = 1$), the model can be re-written as:

$$
\begin{aligned}
Y_1 - \mu_1 &= \theta + \varepsilon_1 \\
\frac{Y_2 - \mu_2}{\alpha_2} &= \theta + \frac{\varepsilon_2}{\alpha_2} \\
\frac{Y_3 - \mu_3}{\alpha_3} &= \theta + \frac{\varepsilon_3}{\alpha_3}
\end{aligned}
$$

and Kotlarski's theorem can be applied to each pair of equations to nonparametrically identify the distributions of θ, ε_1, ε_2/α_2 and ε_3/α_3.

Estimation / Inference

Different approaches:

- **Frequentist:** Maximum likelihood
- **Bayesian:** Markov chain Monte Carlo (MCMC) simulation methods

Estimation of the standard factor model is straightforward to implement, since for individual $i = 1, ..., n$:

$$y_i \mid x_i \sim N(x_i\beta, \Sigma_y) \qquad \Sigma_y = \alpha\Psi\alpha' + \Sigma_\varepsilon$$

The corresponding **individual likelihood** is:

$$L_i(\beta, \alpha, \Psi, \Sigma_\varepsilon; y_i, x_i) = (2\pi)^{-p/2} \mid \Sigma_y \mid^{-1/2} \exp\left\{-\frac{1}{2}(y_i - x_i\beta)'\Sigma_y^{-1}(y_i - x_i\beta)\right\}$$

where p is the number of (continuous) response variables.

For **non-continuous** variables, there is no closed-form solution for the likelihood function:

$$L_i(\beta, \alpha, \Psi, \Sigma_\varepsilon; y_i, x_i) = \int_\Theta p(y_i \mid \theta_i, x_i, \beta, \alpha, \Psi, \Sigma_\varepsilon) p(\theta_i) d\theta_i$$

and **numerical integration** is required.

$\Longrightarrow$ Optimization can be tricky in high-dimensional cases.

Prior beliefs about the distribution of the parameters of interest are updated in light of the data to produce the posterior distribution:

$$\text{prior} \qquad \Longrightarrow \qquad \text{data} \qquad \Longrightarrow \qquad \text{posterior}$$

Bayes' rule to derive the posterior distribution:

$$\underbrace{p(\psi \mid y, x)}_{\text{posterior}} = \frac{p(y \mid \psi, x)p(\psi)}{p(y \mid x)} \propto \underbrace{p(y \mid \psi, x)}_{\text{likelihood}}\underbrace{p(\psi)}_{\text{prior}}$$

where $\psi = (\beta, \alpha, \Psi, \Sigma_\varepsilon)$ represents the set of parameters to be estimated.

The **Gibbs sampler** is widely used to estimate this kind of model [Casella & George, 1992]:

- **Idea:** complicated joint posterior distribution of the parameters is approximated by sequentially drawing from their conditional distributions:

$$p(A, B) \quad \Leftarrow \quad \begin{cases} p(A \mid B) \\ p(B \mid A) \end{cases}$$

- **Data augmentation** procedures [Tanner & Wong, 1987] allow to sample the latent factors and latent response variables.

- **Initialize** the model with starting values for the parameters and the latent factors:

$$\beta^{(0)}, \alpha^{(0)}, \Sigma_\varepsilon^{(0)}, \theta^{(0)}, \Psi^{(0)}$$

- At each iteration (t), **update** model parameters **sequentially** as follows:
  - Sample $\beta^{(t)}$ from $p(\beta \mid y, x, \theta^{(t-1)}, \alpha^{(t-1)}, \Psi^{(t-1)}, \Sigma_\varepsilon^{(t-1)})$
  - Sample $\alpha^{(t)}$ from $p(\alpha \mid y, x, \theta^{(t-1)}, \beta^{(t)}, \Psi^{(t-1)}, \Sigma_\varepsilon^{(t-1)})$
  - Sample $\Sigma_\varepsilon^{(t)}$ from $p(\Sigma_\varepsilon \mid y, x, \theta^{(t-1)}, \beta^{(t)}, \alpha^{(t)}, \Psi^{(t-1)})$
  - Sample $\theta^{(t)}$ from $p(\theta \mid y, x, \beta^{(t)}, \alpha^{(t)}, \Psi^{(t-1)}, \Sigma_\varepsilon^{(t)})$
  - Sample $\Psi^{(t)}$ from $p(\Psi \mid y, x, \theta^{(t)}, \beta^{(t)}, \alpha^{(t)}, \Sigma_\varepsilon^{(t)})$
- Repeat the sampling process **until practical convergence**.

Dynamic factor models

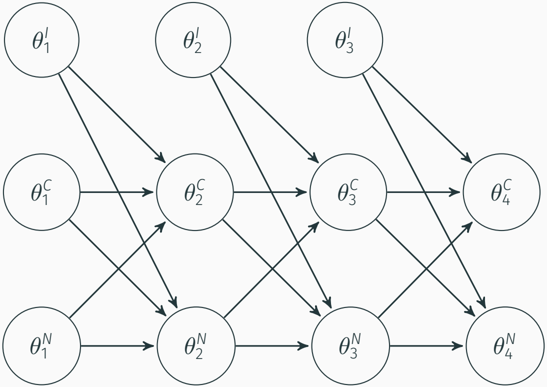
- **So far:** Static factor model for cross-sectional data.
- **Dynamic case:** Repeated measurements over several time periods make it possible to model the evolution and the intertemporal interactions of the latent factors.

Cunha & Heckman (2008) specify a dynamic factor model for the **technology of skill formation**.

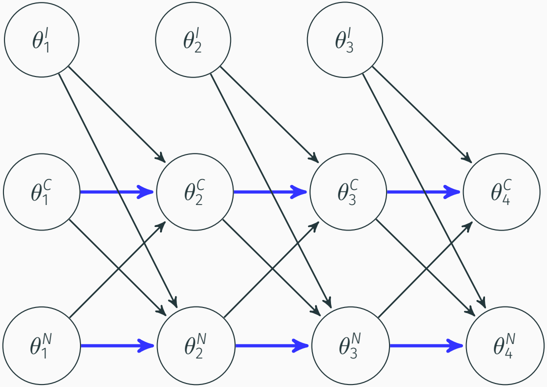
At each period t, the following latent factors are extracted:

- θ_t^C: Cognitive factor
- θ_t^N: Noncognitive factor
- θ_t^I: Parental investment factor

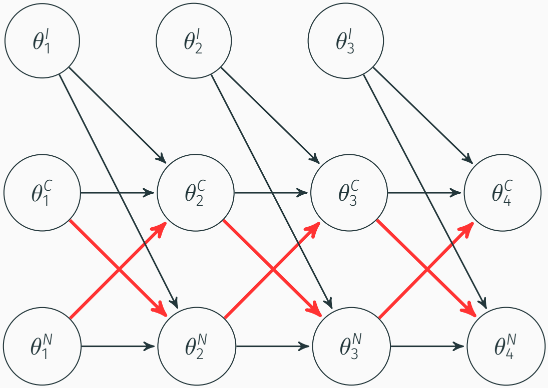
and their interactions over time are captured by the model.


θ_1^I
θ_2^I
θ_3^I
θ_1^C
θ_2^C
θ_3^C
θ_4^C
θ_1^N
θ_2^N
θ_3^N
θ_4^N

Technology of skill formation [Cunha & Heckman, 2008]



$\Longrightarrow$ Self-productivity



$\Longrightarrow$ Cross-productivity

- **Measurement system.** Each latent factor $k \in \{C, N, I\}$, at each period $t = 1, ..., T$, is measured by a set of dedicated items $j = 1, ..., m_t^k$:

$$Y_{jt}^k = \mu_{jt}^k + \alpha_{jt}^k \theta_t^k + \varepsilon_{jt}^k$$

- **Law of motion for skills** (production function for $k \in \{C, N\}$):

$$\theta_{t+1}^k = \gamma_1^k \theta_t^N + \gamma_2^k \theta_t^C + \gamma_3^k \theta_t^I + \eta_t^k$$

  In compact form:

$$\begin{pmatrix} \theta_{t+1}^N \\ \theta_{t+1}^C \end{pmatrix} = \begin{pmatrix} \gamma_1^N & \gamma_2^N \\ \gamma_1^C & \gamma_2^C \end{pmatrix} \begin{pmatrix} \theta_t^N \\ \theta_t^C \end{pmatrix} + \begin{pmatrix} \gamma_3^N \\ \gamma_3^C \end{pmatrix} \theta_t^I + \begin{pmatrix} \eta_t^N \\ \eta_t^C \end{pmatrix}$$

- **Independence assumptions:** different cases of classical and non-classical measurement error are considered in the paper.

- **Self-productivity** of skills:

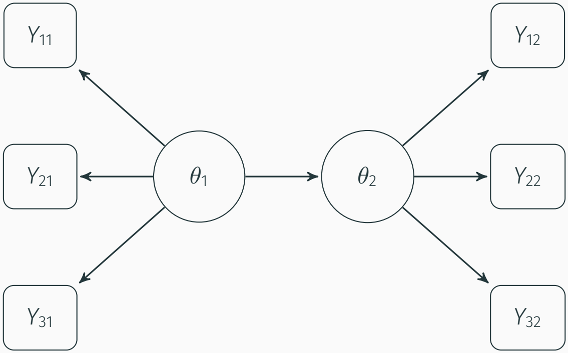
$$\frac{\partial \theta^N_{t+1}}{\partial \theta^N_t} = \gamma^N_1 \qquad \frac{\partial \theta^C_{t+1}}{\partial \theta^C_t} = \gamma^C_2$$

- **Cross-productivity** of skills:

$$\frac{\partial \theta^N_{t+1}}{\partial \theta^C_t} = \gamma^N_2 \qquad \frac{\partial \theta^C_{t+1}}{\partial \theta^N_t} = \gamma^C_1$$

How to identify a dynamic factor model? What restrictions are required?

Consider a very basic example: two-period model, one latent factor, three measurements at each period.


Y_{11}
Y_{21}
Y_{31}
θ_1
θ_2
Y_{12}
Y_{22}
Y_{32}

- Measurement system, for $j = 1, 2, 3$ and $t = 1, 2$:

$$Y_{jt} = \mu_{jt} + \alpha_{jt}\theta_t + \varepsilon_{jt}$$

- Structural part:

$$\theta_2 = \delta\theta_1 + \eta$$

- Independence assumption:

$$\varepsilon_{jt} \perp\!\!\!\perp \varepsilon_{j't'} \perp\!\!\!\perp \theta \qquad \text{and} \qquad \varepsilon_{jt} \perp\!\!\!\perp \eta \perp\!\!\!\perp \theta_1$$

- Distributional assumptions:

$$\varepsilon_{jt} \sim N\left(0, \sigma^2_{jt}\right) \qquad \eta \sim N\left(0, \sigma^2_{\eta}\right)$$

$$\theta_1 \sim N\left(0, \sigma^2_{\theta_1}\right) \qquad \Longrightarrow \theta_2 \sim N\left(0, \delta^2\sigma^2_{\theta_1} + \sigma^2_{\eta}\right)$$

A simple example: identification of the factor loadings

$$cov(Y_{11}, Y_{12}) = \alpha_{11}\alpha_{12}cov(\theta_1, \theta_2)$$
$$cov(Y_{12}, Y_{21}) = \alpha_{12}\alpha_{21}cov(\theta_1, \theta_2)$$
$$cov(Y_{11}, Y_{22}) = \alpha_{11}\alpha_{22}cov(\theta_1, \theta_2)$$

$$cov(Y_{11}, Y_{12}) = \alpha_{11}\alpha_{12}cov(\theta_1, \theta_2)$$
$$cov(Y_{12}, Y_{21}) = \alpha_{12}\alpha_{21}cov(\theta_1, \theta_2)$$
$$cov(Y_{11}, Y_{22}) = \alpha_{11}\alpha_{22}cov(\theta_1, \theta_2)$$

Normalization for identification: $\alpha_{1t} = 1$ for $t = 1, 2$.

$$cov(Y_{11}, Y_{12}) = \alpha_{11}\alpha_{12} cov(\theta_1, \theta_2)$$
$$cov(Y_{12}, Y_{21}) = \alpha_{12}\alpha_{21} cov(\theta_1, \theta_2)$$
$$cov(Y_{11}, Y_{22}) = \alpha_{11}\alpha_{22} cov(\theta_1, \theta_2)$$

Normalization for identification: $\alpha_{1t} = 1$ for $t = 1, 2$.

Taking the following ratios provides:

$$\frac{cov(Y_{12}, Y_{21})}{cov(Y_{11}, Y_{12})} = \alpha_{21} \qquad \frac{cov(Y_{11}, Y_{22})}{cov(Y_{11}, Y_{12})} = \alpha_{22}$$

Loadings α_{31} and α_{32} are identified in the same way.

$$
\begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \end{pmatrix} = \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{31} \\ \mu_{12} \\ \mu_{22} \\ \mu_{32} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \\ \alpha_{31} & 0 \\ 0 & \alpha_{12} \\ 0 & \alpha_{22} \\ 0 & \alpha_{32} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \end{pmatrix}
$$

$$
\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{\theta_1} & \delta\sigma^2_{\theta_1} \\ \delta\sigma^2_{\theta_1} & \delta^2\sigma^2_{\theta_1} + \sigma^2_{\eta} \end{bmatrix}\right)
$$

$$
\varepsilon \sim N\Big(0, \mathrm{diag}(\sigma^2_{11}, \sigma^2_{21}, \sigma^2_{31}, \sigma^2_{12}, \sigma^2_{22}, \sigma^2_{32})\Big)
$$

Error terms ε are assumed to be mutually independent and independent of the latent factors.

This simple two-period dynamic factor model can be re-expressed as a static model with correlated factors, and the standard normalization can be applied for identification purpose:

$$\alpha_{1t} = 1 \qquad \text{for } t = 1, 2$$

to prevent rotation of the factor loading matrix.

References

References

- Geweke & Zhou (1996): “Measuring the Pricing Error of the Arbitrage Pricing Theory,” Review of Financial Studies.
- Früwirth-Schnatter & Lopes (2010): “Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown,” working paper.
- Heckman, Stixrud & Urzua (2006): ”The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” Journal of Labor Economics.
- Kaiser (1958): “The Varimax Criterion for Analytic Rotation in Factor Analysis,” Psychometrika.
- Lopes & West (2004): “Bayesian Model Assessment in Factor Analysis,” Statistica Sinica.
- Rosenberg (1965): *Society and the adolescent self-image*, Princeton University Press.
- Spearman (1904): “General Intelligence, Objectively Determined and Measured,” American Journal of Psychology.

- Casella, G. and George, E.I. (1992): "Explaining the Gibbs Sampler," American Statistician.
- Cunha, F. and J.J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," Journal of Human Resources.
- Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution," Pacific Journal of Mathematics.
- Tanner, M.A. and Wong, W.H. (1987): "The Calculation of Posterior Distributions by Data Augmentation," Journal of the American statistical Association.