How to estimate factor models?

How to estimate factor models?

- Typically, we impose distributional assumptions on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$

How to estimate factor models?

- Typically, we impose distributional assumptions on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$
- e.g. assume $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are each MVN with 0 covariance and $\boldsymbol{\theta} \perp \boldsymbol{\varepsilon}$

How to estimate factor models?

- Typically, we impose distributional assumptions on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$
- e.g. assume $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are each MVN with 0 covariance and $\boldsymbol{\theta} \perp \boldsymbol{\varepsilon}$
- Then we estimate $(\boldsymbol{\beta}, \Lambda, \Sigma_{\boldsymbol{\theta}}, \Omega)$ by maximum likelihood

How to estimate factor models?

- Typically, we impose distributional assumptions on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$
- e.g. assume $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are each MVN with 0 covariance and $\boldsymbol{\theta} \perp \boldsymbol{\varepsilon}$
- Then we estimate $(\boldsymbol{\beta}, \Lambda, \Sigma_{\boldsymbol{\theta}}, \Omega)$ by maximum likelihood
- The likelihood function will need to be integrated, since $\boldsymbol{\theta}$ is unobserved

How to estimate factor models?

- Typically, we impose distributional assumptions on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$
- e.g. assume $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are each MVN with 0 covariance and $\boldsymbol{\theta} \perp \boldsymbol{\varepsilon}$
- Then we estimate $(\boldsymbol{\beta}, \Lambda, \Sigma_{\boldsymbol{\theta}}, \Omega)$ by maximum likelihood
- The likelihood function will need to be integrated, since $\boldsymbol{\theta}$ is unobserved
- Can use quadrature, MC integration, SMM, MCMC, or EM / MM algorithms

How to estimate factor models?

- Typically, we impose distributional assumptions on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$
- e.g. assume $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$ are each MVN with 0 covariance and $\boldsymbol{\theta} \perp \boldsymbol{\varepsilon}$
- Then we estimate $(\boldsymbol{\beta}, \Lambda, \Sigma_{\boldsymbol{\theta}}, \Omega)$ by maximum likelihood
- The likelihood function will need to be integrated, since $\boldsymbol{\theta}$ is unobserved
- Can use quadrature, MC integration, SMM, MCMC, or EM / MM algorithms
- As you know, these vary in their ease of use!

- The whole reason we use a factor model is to reduce bias and improve precision

- The whole reason we use a factor model is to reduce bias and improve precision

- Let's go back to the log wage example from earlier

- The whole reason we use a factor model is to reduce bias and improve precision
- Let's go back to the log wage example from earlier
    - β's are biased if we omit cognitive ability (omitted variable bias)

- The whole reason we use a factor model is to reduce bias and improve precision
- Let's go back to the log wage example from earlier
  - β's are biased if we omit cognitive ability (omitted variable bias)
  - β's are also biased if we include IQ score (attenuation bias from meas. err.)

- The whole reason we use a factor model is to reduce bias and improve precision
- Let's go back to the log wage example from earlier
  - β's are biased if we omit cognitive ability (omitted variable bias)
  - β's are also biased if we include IQ score (attenuation bias from meas. err.)
  - We know ability affects wages, and we have (noisy & correlated) measurements of it

- The whole reason we use a factor model is to reduce bias and improve precision
- Let's go back to the log wage example from earlier
  - β's are biased if we omit cognitive ability (omitted variable bias)
  - β's are also biased if we include IQ score (attenuation bias from meas. err.)
  - We know ability affects wages, and we have (noisy & correlated) measurements of it
- We can estimate the log wage parameters by maximum likelihood

- The whole reason we use a factor model is to reduce bias and improve precision
- Let's go back to the log wage example from earlier
    - β's are biased if we omit cognitive ability (omitted variable bias)
    - β's are also biased if we include IQ score (attenuation bias from meas. err.)
    - We know ability affects wages, and we have (noisy & correlated) measurements of it
- We can estimate the log wage parameters by maximum likelihood
- We combine together the log wage and factor model likelihoods

- The whole reason we use a factor model is to reduce bias and improve precision
- Let's go back to the log wage example from earlier
  - β's are biased if we omit cognitive ability (omitted variable bias)
  - β's are also biased if we include IQ score (attenuation bias from meas. err.)
  - We know ability affects wages, and we have (noisy & correlated) measurements of it
- We can estimate the log wage parameters by maximum likelihood
- We combine together the log wage and factor model likelihoods
- Coding walkthrough later

Log wage example: system of equations

Log wage example: system of equations

$$
\begin{aligned}
asvab_j &= W\alpha_j + \gamma_j \xi + \varepsilon_j \\
\log(wage) &= X\beta + \delta\xi + \eta
\end{aligned}
$$

where $\xi \sim N(0,1)$ is unobserved cognitive ability

Log wage example: system of equations

$$\begin{aligned} asvab_j &= W\alpha_j + \gamma_j \xi + \varepsilon_j \\ \log(wage) &= X\beta + \delta\xi + \eta \end{aligned}$$

where $\xi \sim N(0,1)$ is unobserved cognitive ability

Likelihood for observation i, assuming $\eta \sim N(0,\sigma_w^2)$ and $\varepsilon_j \sim N(0,\sigma_j^2)$:

$$\mathcal{L}_i = \left\{ \prod_j \frac{1}{\sigma_j} \phi\left(\frac{asvab_{ij} - W_i\alpha_j - \gamma_j \xi_i}{\sigma_j} \right) \right\} \frac{1}{\sigma_w} \phi\left(\frac{\log(wage_i) - X_i\beta - \delta\xi_i}{\sigma_w} \right)$$

Since ξ is unobserved, integrate out: $\ell = \sum_i \log\left(\int \mathcal{L}_i \mathrm{d}F(\xi)\right)$

- Factor models can also be used to account for dynamic selection

- Factor models can also be used to account for dynamic selection

- Intuition: $\uparrow$ cog. abil $\Rightarrow$ $\uparrow$ schooling $\Rightarrow$ $\uparrow$ wages

- Factor models can also be used to account for dynamic selection
- Intuition: $\uparrow$ cog. abil $\Rightarrow$ $\uparrow$ schooling $\Rightarrow$ $\uparrow$ wages
- Schooling is endogenous, so we can add a schooling choice model to our likelihood

- Factor models can also be used to account for dynamic selection
- Intuition: $\uparrow$ cog. abil $\Rightarrow$ $\uparrow$ schooling $\Rightarrow$ $\uparrow$ wages
- Schooling is endogenous, so we can add a schooling choice model to our likelihood
- Ability factor $\rightarrow$ choice of schooling $\Rightarrow$ corr. between schooling choices and wages

- Factor models can also be used to account for dynamic selection

- Intuition: $\uparrow$ cog. abil $\Rightarrow$ $\uparrow$ schooling $\Rightarrow$ $\uparrow$ wages

- Schooling is endogenous, so we can add a schooling choice model to our likelihood

- Ability factor $\rightarrow$ choice of schooling $\Rightarrow$ corr. between schooling choices and wages

- But conditional on the factor, we have separability of the likelihood components

- Factor models can also be used to account for dynamic selection

- Intuition: $\uparrow$ cog. abil $\Rightarrow \uparrow$ schooling $\Rightarrow \uparrow$ wages

- Schooling is endogenous, so we can add a schooling choice model to our likelihood

- Ability factor $\rightarrow$ choice of schooling $\Rightarrow$ corr. between schooling choices and wages

- But conditional on the factor, we have separability of the likelihood components

$$\mathcal{L} = \int \underbrace{\mathcal{L}_1(A)}_{\text{measurements}} \underbrace{\mathcal{L}_2(A)}_{\text{choices}} \underbrace{\mathcal{L}_3(A)}_{\text{wages}} dF(A)$$

- Factor models can also be used to account for dynamic selection

- Intuition: $\uparrow$ cog. abil $\Rightarrow$ $\uparrow$ schooling $\Rightarrow$ $\uparrow$ wages

- Schooling is endogenous, so we can add a schooling choice model to our likelihood

- Ability factor $\rightarrow$ choice of schooling $\Rightarrow$ corr. between schooling choices and wages

- But conditional on the factor, we have separability of the likelihood components

$$\mathcal{L} = \int \underbrace{\mathcal{L}_1(A)}_{\text{measurements}} \underbrace{\mathcal{L}_2(A)}_{\text{choices}} \underbrace{\mathcal{L}_3(A)}_{\text{wages}} dF(A)$$

- Reminiscent of mixed logit / preference heterogeneity