Generalized Linear Factor Models: a local EM estimation

Xavier Bry [a], ***Christian Lavergne*** [ab] and ***Mohamed Saidane*** [c]

E-mails: [bry , lavergne]@math.univ-montp2.fr ; Mohamed.Saidane@isg.rnu.tn

a I3M, Université Montpellier II, France
b université Montpellier III, France
c Université du 7 Novembre à Carthage, Tunisie

Motivations

Motivations:

Social Sciences, Biology, Environment, ... → Quantitative measures / Qualitative characteristics / Counts / Life times = Miscellaneous types of variables

$$B(1, p)\,;\, B(n\,,\, p)$$
$$M\,(1\,;\, p_1\,,\, p_k)\,;\, M\,(n\,;\, p_1\,,\, p_k)$$
$$P(\lambda)\,;\, \gamma\,(k\,,\lambda)\,;\, etc.$$

Motivations

Motivations:

Social Sciences, Biology, Environment, ... → Quantitative measures, Qualitative characteristics, Counts, Life times = Miscellaneous types of variables

$$B(1,p)\,;\,B(n,p)$$
$$M(1\,;\,p_1,\,p_k)\,;\,M(n\,;\,p_1,\,p_k)$$
$$P(\lambda)\,;\,\gamma(k,\lambda)\,;\,etc.$$

Abundant & related

⇒ Dimension reduction

⇒ Correlation modeling

Model & notations

Data:

unit t

Observed on n observation units $\{1, \ldots, t, \ldots n\}$: p variables $\{y_1, \ldots, y_p\}$ $\rightarrow$ $\mathrm{y}_t = (y_{it})_{i=1,p}$ $(p,1)$

underlying

q latent factors $\{f_1, \ldots, f_q\}$ $\rightarrow$ $\mathrm{f}_t = (f_{jt})_{j=1,q}$ $(q,1)$

$q < p$

Observation units are independent

Model & notations

unit t ↓

Data:

Observed on n observation units $\{1, \ldots, t, \ldots n\}$: p variables $\{y_1, \ldots, y_p\}$ → $\mathrm{y}_t = (y_{it})_{i=1,p}$ $(p,1)$

underlying

q latent factors $\{f_1, \ldots, f_q\}$ → $\mathrm{f}_t = (f_{jt})_{j=1,q}$ $(q,1)$

$q < p$

Observation units are independent

Factor Model:

$\forall t,\ \mathrm{f}_t \sim N(0\,;I_q)$

Factors f_j generate linear predictors of variables y_i

→ linear predictor of $y_{it} \mid \mathrm{f}_t$: $\eta_{it} = \theta_i + a_i'\mathrm{f}_t$ → $\underset{(p,1)}{\eta_t} = \underset{(p,1)}{\theta} + \underset{(p,q)}{A}\,\underset{(q,1)}{\mathrm{f}_t}$

$A = (a_1, \ldots, a_p)'$; $\theta = (\theta_i)_i$; $\mathrm{F} = (\mathrm{f}_1, \ldots, \mathrm{f}_t, \ldots, \mathrm{f}_n)$;

$\eta_t = (\eta_{it})_i$; $\eta = (\eta_{it})_{i,t} = (\eta_1, \ldots, \eta_t, \ldots, \eta_n)$

$$\underset{(p,n)}{\eta} = \theta\,\mathbf{1}_n' + \underset{(p,q)(q,n)}{A\,F}$$

Model & notations

Model of y conditional to F:

$\forall t,\ \mathrm{y}_t \mid \mathrm{f}_t \ \sim\ \wp_t \in$ Exponential family (Nelder & Wedderburn): $l_i(y_{it}|\delta_{it},\phi)=\exp\left(\dfrac{y_{it}\delta_{it}-b_i(\delta_{it})}{a_{it}(\phi)}+c_i(y_{it},\phi)\right)$

$\forall t,\ (y_{it})_i \mid \mathrm{f}_t$ are $\perp\!\!\!\perp$

$\mu_{it} = E(y_{it}) = b_i'(\delta_{it})$

$Var(y_{it}) = a_{it}(\phi)\, b_i''(\delta_{it})$

$= a_{it}(\phi)\, \underbrace{b_i''(b_i'^{-1}(\mu_{it}))}$

$v_i(\mu_{it}) = b_i''([b_i'^{-1}(\mu_{it})]$

Conditional variance matrix: $Var(y_t)=diag\{a_{it}(\phi)v_i(\mu_{it})\}_{i=1,p}$

Model & notations

Model of y conditional to F:

$\forall t,\ \mathrm{y}_t \mid \mathrm{f}_t \sim \wp_t \in$ Exponential family (Nelder & Wedderburn): $l_i(y_{it} | \delta_{it}, \phi) = \exp\left(\dfrac{y_{it}\delta_{it} - b_i(\delta_{it})}{a_{it}(\phi)} + c_i(y_{it}, \phi) \right)$

$\forall t,\ (y_{it})_i \mid \mathrm{f}_t$ are $\perp\!\!\!\perp$

$\mu_{it} = E(y_{it}) = b_i'(\delta_{it})$

$$Var(y_{it}) = a_{it}(\phi)\, b_i''(\delta_{it}) = a_{it}(\phi)\, \underbrace{b_i''(b_i'^{-1}(\mu_{it}))}_{v_i(\mu_{it}) = b_i''([b_i'^{-1}(\mu_{it})]}$$

Conditional variance matrix: $Var(y_t) = diag\{a_{it}(\phi)\, v_i(\mu_{it})\}_{i=1,p}$

Link with linear predictor: $\forall\ i, t:\ \eta_{it} = \underbrace{g_i}_{\textit{link function}}(\mu_{it})$

δ_{it} = canonical parameter ; $\underbrace{g_i = b_i'^{-1}}_{\textit{canonical link}} \Rightarrow \eta_{it} = \delta_{it}$

The classical Gaussian Linear Factor Model: $\mathrm{y}_{it} \mid \mathrm{f}_t \sim N(\mu_{it}; \sigma^2)$ with $\mu_{it} = \eta_{it}$

Factor Models: available estimation techniques

The classical Gaussian Linear Factor Model:

EM algorithm estimation:

A is estimated by maximizing the expectation, conditional to observations, of the derivative of the completed log-likelihood (EDLCO), integrated with respect to the factors:

$\Rightarrow$ taking the conditional expectation of the 1st order conditions: $\sum_t E\left(\nabla \log l(y_t, f_t) \middle| y_t\right) = 0$

= possible because EDLCO is analytically determined

Factor Models: available estimation techniques

The classical Gaussian Linear Factor Model:

EM algorithm estimation:

A is estimated by maximizing the expectation, conditional to observations, of the derivative of the completed log-likelihood (EDLCO), integrated with respect to the factors:

$\Rightarrow$ taking the conditional expectation of the 1st order conditions: $\sum_t E\left(\nabla \log l(y_t, f_t) | y_t\right) = 0$

= possible because EDLCO is analytically determined

Generalized Linear Factor Models:

Expectation of EDLCO not analytically determined → Direct EM impossible

Computa-tionally intensive:

[Moustaki, I., & Knott, M. (2000)] → Max of the expected completed log- Likelihood, single factor; Gauss-Hermite quadrature used to approximate integral.

[Wedel, M. & Kamakura, W.A. (2001)] → Monte Carlo approach

[Moustaki, I & Victoria-Feser, M.P.(2006)] → Iterative estimation method inspired from the indirect inference technique [Gourieroux (1993)].

Looking back at GLM's

GLM:

GLM of one variable y, depending on predictors $X = (x_j)_{j=1,q}$

$\mu = E(y)$

Linear predictor: $\eta = X\beta$ $\quad \forall t,\ \eta_t = g(\mu_t) \Rightarrow x_t'\beta = g(b'(\delta_t))$

Log-likelihood: $$L(\delta\,;y)=\sum_{t=1}^{n} L_t(\delta_t;y_t)=\sum_{t=1}^{n}\left(\frac{y_t\delta_t-b(\delta_t)}{a_t(\phi)}+c(y_t,\phi)\right)$$

Derivation / $\boldsymbol{\beta}$: $$\frac{\partial L_t}{\partial\beta_j}=\frac{\partial\eta_t}{\partial\beta_j}\frac{\partial\mu_t}{\partial\eta_t}\frac{\partial\delta_t}{\partial\mu_t}\frac{\partial L_t}{\partial\delta_t}=x_{tj}\frac{1}{g'(\mu_t)}\frac{1}{b''(\delta_t)}\frac{y_t-\mu_t}{a_t(\phi)} \longrightarrow = V(y_t) = a_t(\phi)v(\mu_t)$$

Let: $$W_\beta=diag\left(g'(\mu_t)^2 V(y_t)\right)_{t=1,n}=diag\left(g'(\mu_t)^2 a_t(\phi)v(\mu_t)\right)_{t=1,n}\ ;$$

$$\frac{\partial\eta}{\partial\mu}=diag\left(\frac{\partial\eta_t}{\partial\mu_t}\right)_{t=1,n}=diag\left(g'(\mu_t)\right)_{t=1,n}$$

Then: $$\nabla_\beta L=0 \Leftrightarrow X'W_\beta^{-1}\frac{\partial\eta}{\partial\mu}(y-\mu)=0$$ non linear / β

interpretable as normal equations of a linear model

Looking back at GLM's

Fisher's scores algorithm:

$$\beta^{[k+1]}=\underset{\substack{\uparrow\\ \textit{iteration nr.}}}{\beta^{[k]}}-\left(E\left[\frac{\partial^2 L}{\partial\beta\,\partial\beta'}\right]^{[k]}\right)^{-1}\left(\frac{\partial L}{\partial\beta}\right)^{[k]} \quad = \quad \beta^{[k]}-\left(X'W^{-1}_{\beta^{[k]}}X\right)^{-1}X'W^{-1}_{\beta^{[k]}}\left(\frac{\partial\eta}{\partial\mu}\right)^{[k]}(y-\mu^{[k]})$$

$$=\left(X'W^{-1}_{\beta^{[k]}}X\right)^{-1}X'W^{-1}_{\beta^{[k]}}\underbrace{\left(X\beta^{[k]}+\left(\frac{\partial\eta}{\partial\mu}\right)^{[k]}(y-\mu^{[k]})\right)}_{z^{[k]}\text{ = working variable}}$$

$$\nabla_\beta L=0 \Leftrightarrow X'W_\beta^{-1}\frac{\partial\eta}{\partial\mu}(y-\mu)=0$$

$$\Leftrightarrow X'W_\beta^{-1}(z-X\beta)=0$$

normal equations of lin. model *M*: $z_\beta=X\beta+\zeta \;;\; E(\zeta)=0$

$V(\zeta_t)=V(z_{\beta,t})=g'^2(\mu_t)V(y_t):V(\zeta)=W_\beta$

Current linearized model: $M^{[k]}:\; z_\beta=X\beta+\zeta^{[k]} \;;\; E(\zeta^{[k]})=0\,,V(\zeta^{[k]})=W_{\beta^{[k]}}$

Looking back at GLM's

Fisher's scores algorithm:

$$\beta^{[k+1]} = \underset{\uparrow}{\beta^{[k]}} - \left(E\left[\frac{\partial^2 L}{\partial\beta\,\partial\beta'}\right]^{[k]}\right)^{-1}\left(\frac{\partial L}{\partial\beta}\right)^{[k]} \quad = \quad \beta^{[k]} - \left(X'W_{\beta^{[k]}}^{-1}X\right)^{-1}X'W_{\beta^{[k]}}^{-1}\left(\frac{\partial\eta}{\partial\mu}\right)^{[k]}(y-\mu^{[k]})$$

iteration nr.

$$= \left(X'W_{\beta^{[k]}}^{-1}X\right)^{-1}X'W_{\beta^{[k]}}^{-1}\underbrace{\left(X\beta^{[k]} + \left(\frac{\partial\eta}{\partial\mu}\right)^{[k]}(y-\mu^{[k]})\right)}_{z^{[k]} = \text{working variable}}$$

$$\nabla_\beta L = 0 \;\Leftrightarrow\; X'W_\beta^{-1}\frac{\partial\eta}{\partial\mu}(y-\mu) = 0$$

$$\Leftrightarrow\; X'W_\beta^{-1}(z - X\beta) = 0$$

normal equations of lin. model M: $z_\beta = X\beta + \zeta \;;\; E(\zeta) = 0$

$V(\zeta_t) = V(z_{\beta,t}) = g'^2(\mu_t)V(y_t) : V(\zeta) = W_\beta$

Current linearized model: $M^{[k]}:\; z_\beta = X\beta + \zeta^{[k]} \;;\; E(\zeta^{[k]}) = 0,\, V(\zeta^{[k]}) = W_{\beta^{[k]}}$

Iterative GLS estimation:

0) Initializing $M^{[0]}$ with OLS of $g(y)$ on $X \rightarrow \beta^{[0]}$

$$g(y) \approx g(\mu) + g'(\mu)\,(y-\mu) = X\beta + \left(\frac{\partial\eta}{\partial\mu}\right)(y-\mu) = z$$

i) $\beta^{[k]} \rightarrow W_{\beta^{[k]}} \;;\; z_{\beta^{[k]}}$

ii) GLS on $M^{[k]} \rightarrow \beta^{[k]}$

Repeat until convergence

= Quasi-Likelihood Estimation (QLE) = mimics MLE on each step, under a normality and independence assumption of the $z_{\beta,t}$'s with a fixed covariance structure.

The Local EM algorithm

The idea: GLFM = GLM model conditional to *F*

= FM model within the linearized model locally mimicking this GLM

Local QLE of GLM mimics MLE of a *gaussian linear model*

Gaussian linear factor models may be estimated through EM

EM used on the linearized model in Fisher's scores algorithm

The Local EM algorithm

The idea: GLFM = GLM model conditional to *F*

= FM model within the linearized model locally mimicking this GLM

Local QLE of GLM mimics MLE of a *gaussian linear model*

Gaussian linear factor models may be estimated through EM

EM used on the linearized model in Fisher's scores algorithm

The algorithm:

(i) Conditional to θ, A, F, calculate:

$$\varepsilon_{i,F} = y_i - \mu_{i,F} \qquad \zeta_{i,F} = g'(\mu_{i,F})\varepsilon_{i,F} \qquad z_{i,F} = \theta_i \mathbf{1}_n + F a_i + \zeta_{i,F}$$

(ii) Given Z and $V(\zeta)$, we have the linearized *marginal* model:

$\forall t = 1, n: \ z_t = \theta + A f_t + \zeta_t$ viewed as a non-standard FM estimated through an EM step, yielding F:

$$f_t \sim N(0, I_k) \Rightarrow V(z_t) = \Sigma_t = AA' + \Psi_t \ \text{ with } \ \Psi_t = diag(g'^2(\mu_{i,f_t}) V(\varepsilon_{it}|f_t))_{i=1,p}$$

If g = canonical link: $Var(\varepsilon_{it}) = Var(y_{it}) = a_{it}(\phi)\, b_i''([b_i'^{-1}(\mu_{it})]) = a_{it}(\phi)\, g_i^{-1\prime}([g_i(\mu_{it})]) = a_{it}(\phi) / g_i'(\mu_{it})$

$$\Rightarrow \Psi_t = diag(a_{it}(\phi) g'(\mu_{i,f_t}))_{i=1,p} \qquad \to \text{EM uses } \Sigma_t$$

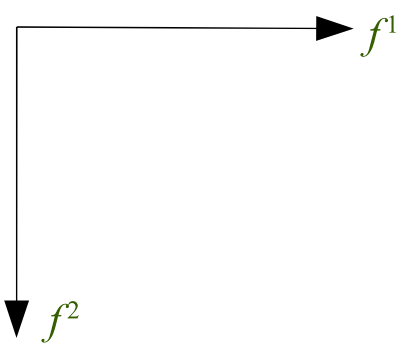
(iii) Given F, we have the linearized *conditional* model, viewed as a GLM → FSA updates θ and A using variance matrix: $V(z_t|F_t) = V(\zeta_t) = \Psi_t$

A quick review of performances on simulated data

Model : Poisson

$\forall t = 1$ to 400 , $\forall i = 1, 40$: $y_{it} \mid \mathrm{f}_t \sim \wp(e^{\eta_{it}})$ independently

Simulation using 2 $\mathcal{N}(0;1)$ factors: $\mathrm{f}_t = (\mathrm{f}_t^1 \, , \, \mathrm{f}_t^2)$: $\eta_{it} = \theta_i + a_{i1}\,\mathrm{f}_t^1 + a_{i2}\,\mathrm{f}_t^2$



A quick review of performances on simulated data

Model : Poisson

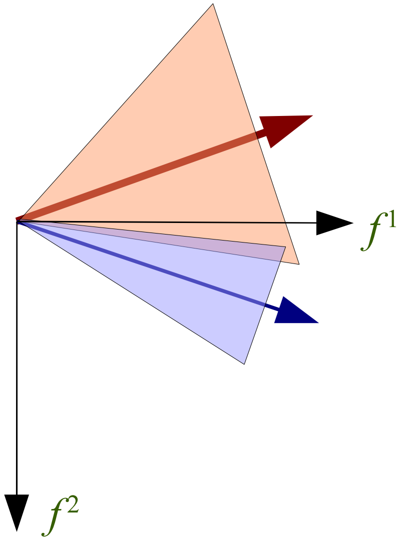
$\forall t = 1$ to 400 , $\forall i = 1, 40$: $y_{it} \mid f_t \sim \wp(e^{\eta_{it}})$ independently

Simulation using 2 $N(0;1)$ factors: $f_t = (f_t^1 , f_t^2)$: $\eta_{it} = \theta_i + a_{i1} f_t^1 + a_{i2} f_t^2$

40 "observed" variables y^i structured in 2 bundles + noise:

Bundle 1: y^1 to y^{25}

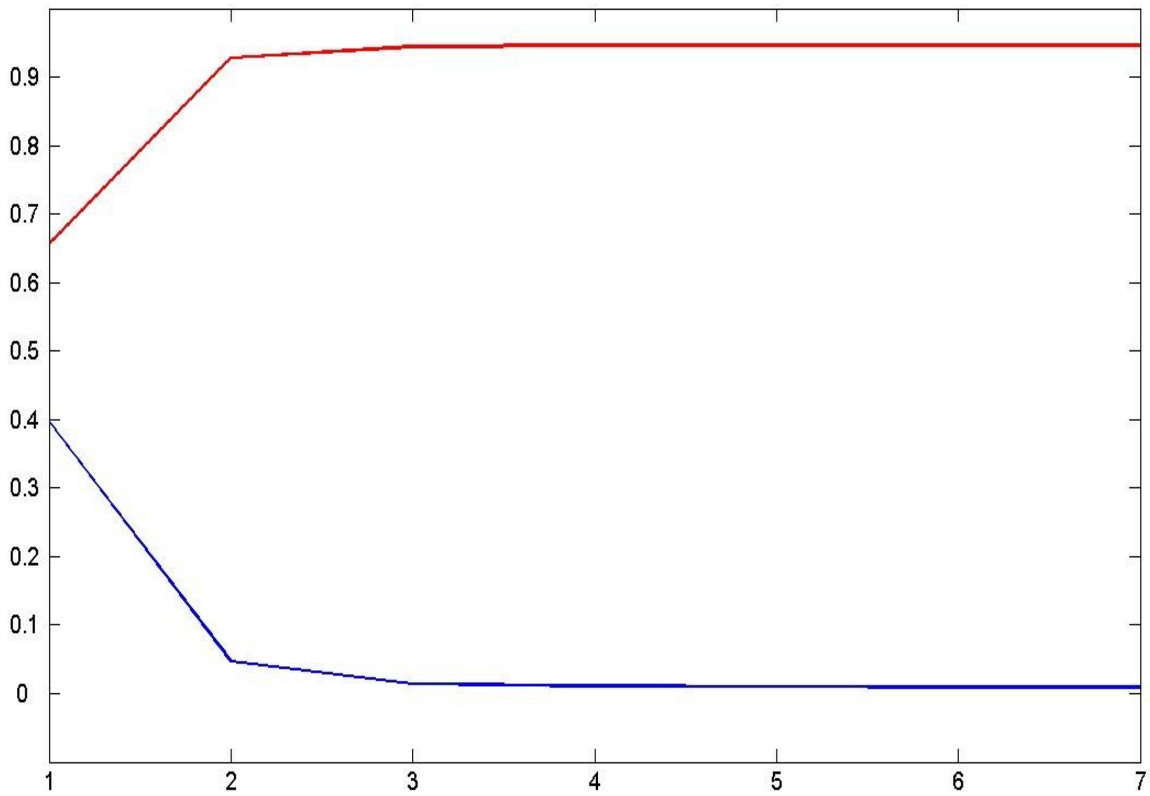
Bundle 2: y^{26} to y^{40}



A quick review of performances on simulated data

Convergence facts:

Average correlation between real factors and estimated ones:

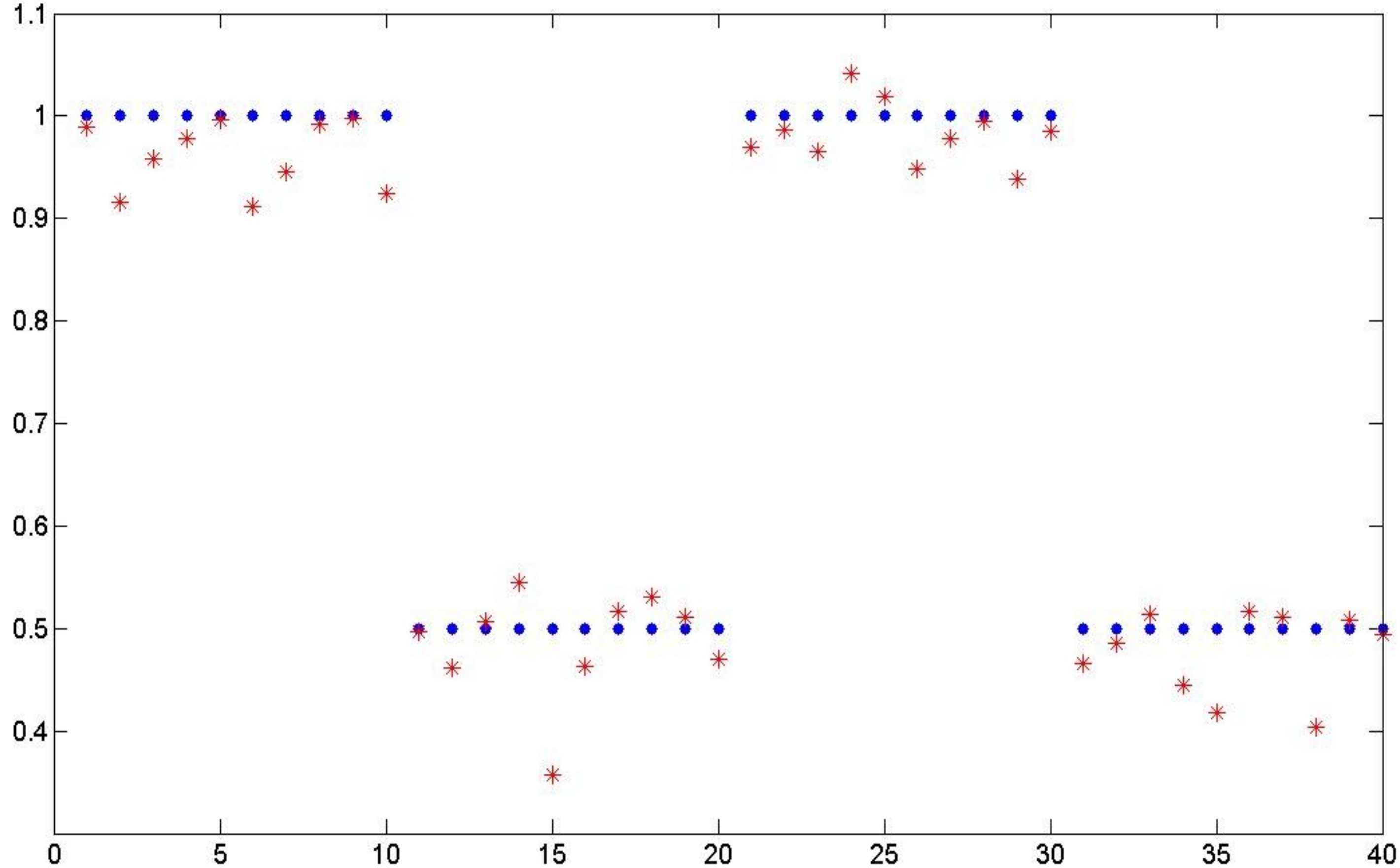


nb of iterations

A quick review of performances on simulated data

Estimation accuracy:

Means: real values* vs *estimates:

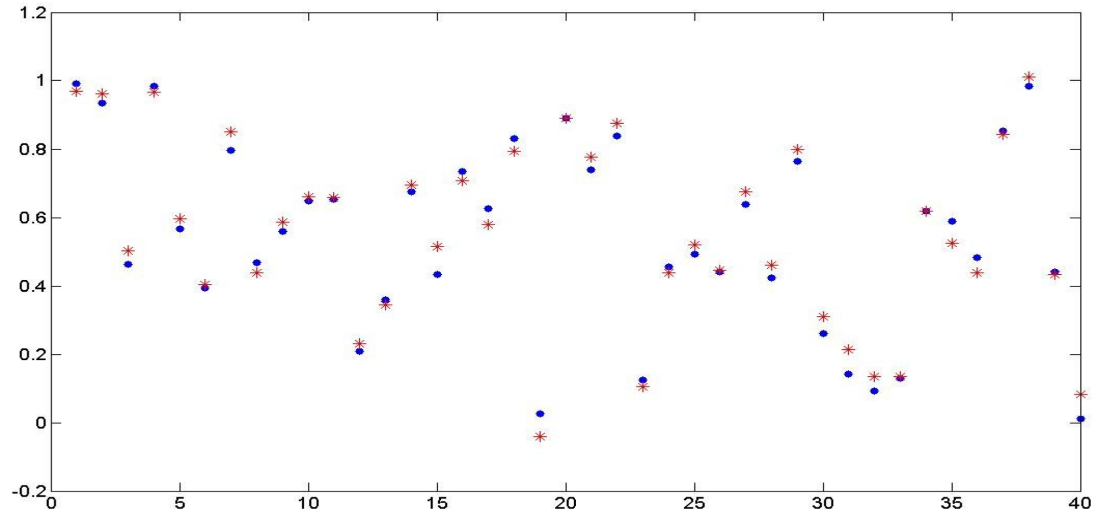


A quick review of performances on simulated data

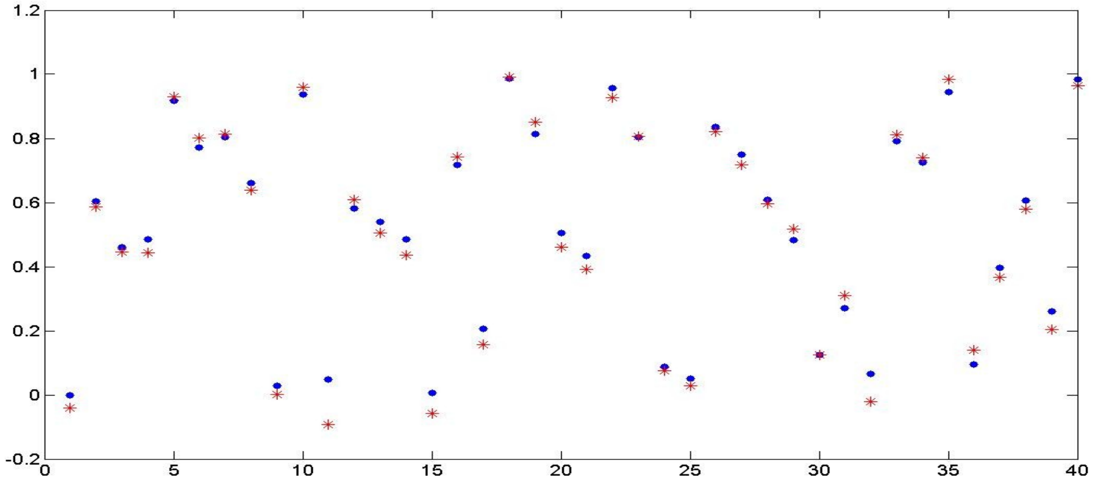
Estimation accuracy:

Coefficients a_{ij}: real values* vs *estimates:

Factor 1:



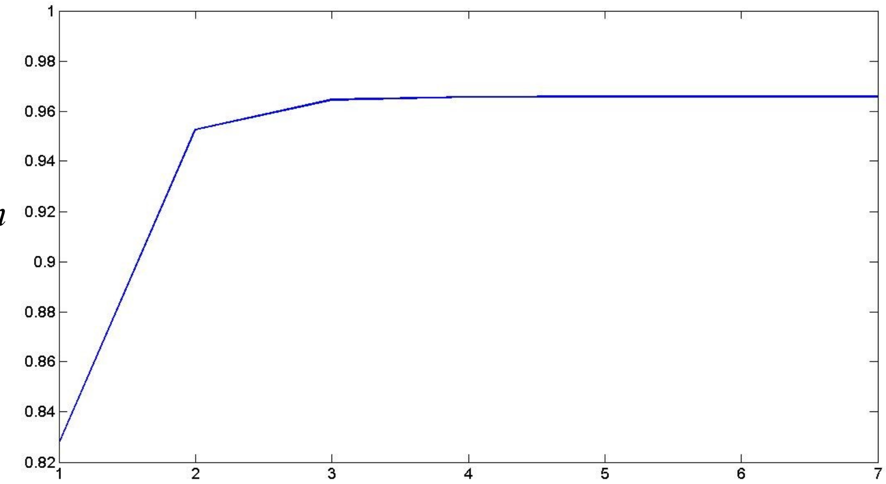
Factor 2:



A quick review of performances on simulated data

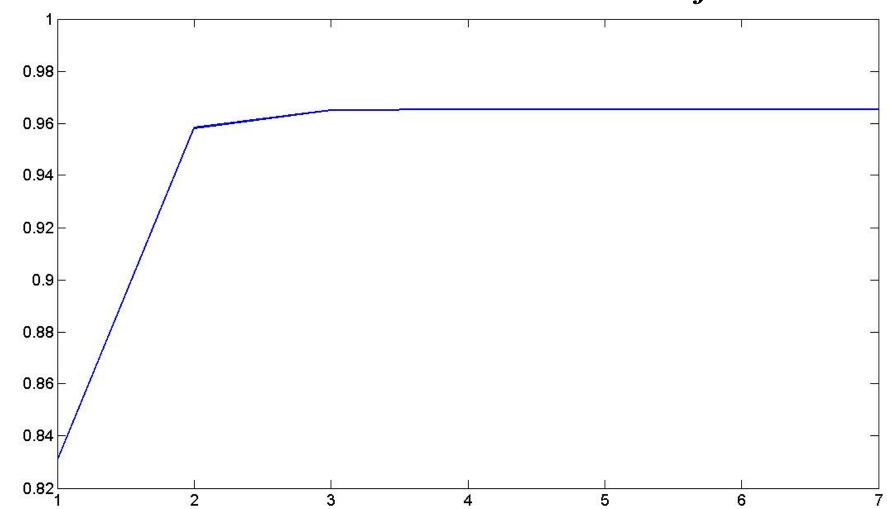
When simulation coefficients do not meet identification constraints...

R^2 of real factor f_1 on estimated $<f_1\ , f_2>$:

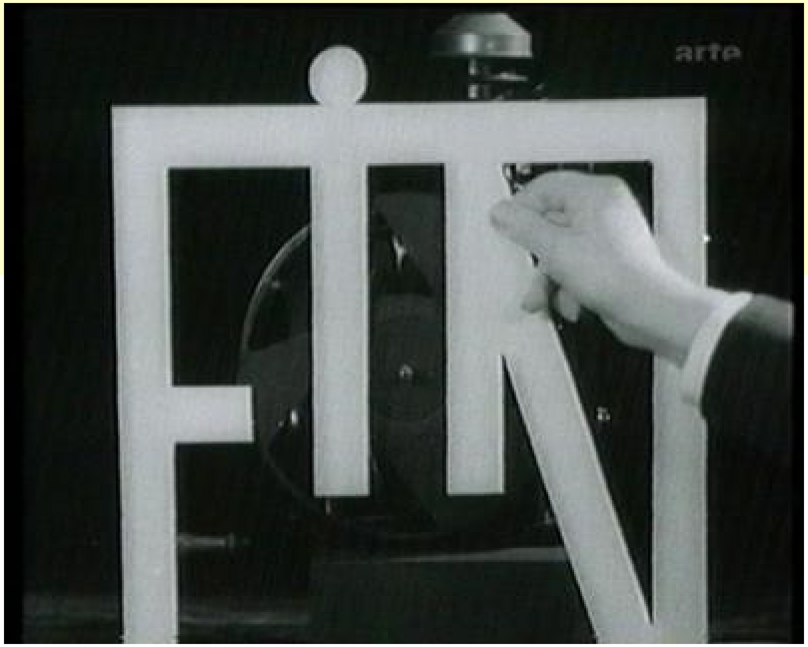


nb of iterations

R^2 of real factor f_2 on estimated $<f_1\ , f_2>$:




The End
THE END
A Universal Release
THE END
The End
The End
FIN


arte
FIN


The end