Contributions to The Estimation of Latent Factorial Generalized Linear Models

Mohamed Saidane, Christian Lavergne, Xavier Bry

Contributions to The Estimation of Latent Factorial Generalized Linear Models

Mohamed Saidane †, Christian Lavergne ‡ and Xavier Bry ‡

† Université de Carthage
ISCC Bizerte, Zarzouna 7021 - Bizerte - Tunisie
E-mail: Mohamed.Saidane@isg.rnu.tn
‡ Institut de Mathématiques et de Modélisation de Montpellier
UMR-CNRS 5149, Université Montpellier II
Place Eugène Bataillon CC 051 – 34095, France
E-mail: [lavergne , bry]@math.univ-montp2.fr

Abstract

Factor models were first developed and dealt with in the case where observations are assumed to be normally distributed. Estimation is then carried out using the Expectation-Maximization (EM) algorithm based on the fact that the expectation of the completed log-likelihood conditional to the data is available in such a case. More recently, a less restrictive framework has been considered, in which the distribution of the observations is assumed to belong to the exponential family. We call these models Generalized Linear Factor Models (GLFM). For want of an explicit expression of the expectation of the completed log-likelihood conditional to the data, estimation of a GLFM is currently carried out using Monte-Carlo methods, which are computationally intensive. Here, we propose a quicker estimation technique, based on the fact that the estimation of a Generalized Linear Model can be achieved using the Fisher's Score Algorithm (FSA), that iterates GLS on a locally linearized model. The linearized model mimicking a classical normal factor model, it can be estimated with the EM algorithm. So, our technique consists in nesting an EM algorithm within each iteration of the FSA. Extensive Monte Carlo simulations show promising results of the algorithm.

AMS Subject Classification: 62H25, 62J12, 62P05 and 62P20

Keywords: Factor Models; Generalized Linear Models; EM Algorithm; Scores Algorithm; Simulations.

1 Introduction

Latent variable models are widely used in social sciences for studying the interrelationships among observed variables. More specifically, latent variable models are used for reducing the dimensionality of multivariate data, for assigning scores to sample members on the

latent dimensions identified by the model, and for constructing measurement scales (e.g., in psychometrics). [10, 11] proposed a generalized linear latent variable model framework for any type of observed data (metric or categorical) in the exponential family. They extended the work of [9] and [16] for mixed binary and metric variables (the latter with covariate effects as well) and [2] for categorical variables. A similar framework was also discussed by [17] that includes multilevel models (random-effects models) as a special case.

In this paper we develop a general approach to factor analysis that involves observed variables that are assumed to be distributed in the exponential family. It accommodates a great variety of data, including rating, ordering, choice, frequency, and timing data and entails a number of special cases of factor analysis not considered previously.

The framework is that of factor models (FM): a set of p observed random variables (RV) $\{y_1, ..., \ y_p\}$ is assumed to be produced by fewer ($q \leq p$) unobserved (latent) ones, $\{f_1, ..., \ f_q\}$, called factors. In the beginning, developments on FM's were limited by the assumption that $\{y_1, ..., \ y_p\}$ were normally distributed, and used this specific distribution to carry out their estimation, through the EM algorithm ([8] and [13, 14]). EM could be used then because within the normal framework, the expectation of the derivative of the completed log-likelihood conditional to the data can be calculated analytically.

Such a classical normal factor model was extended later, by considering y's that had a distribution belonging to the exponential family. In such a case, EM could not be carried out, for want of an analytic expression of the expectation of the above-mentioned derivative. So, this expectation had to be approximated to some extent. [10] used the Gauss-Hermite quadrature to approximate integrals in the expectation. [18] used a Monte Carlo method, namely a simulated EM, to approximate the expectation. Their model considered factors not only normally distributed, but more generally distributed in the exponential family. [11] applied the indirect inference technique proposed by [6], [4] and [5]. All these authors only dealt with a single factor model. Such methods are computationally very intensive, which precludes their use on massive data. As a consequence, a quicker estimation technique has to be used in this latter case.

The paper is organized as follows. In section 2, we give the structure of the GLFM. In section 3, we build up our estimation technique by combining the Fisher's score algorithm with the EM algorithm. In section 4, we finally study the performance of this technique on simulated data.

2 General structure of the Generalized Linear Factor Model

Variables $\{y_1, ..., y_p\}$ are measured out on n independent observation units $\{1, ..., \ t, ... \ n\}$. Conditional to the factors $\{f_1, ..., \ f_q\}$, each y_i is modeled with a GLM taking these factors as predictors. For identification purposes, the factors are taken uncorrelated and normally distributed with 0 mean and unit variance:

2.1 Model of the dependent variable vector conditional to factors

Let $y_t = (y_{it})_{i=1,p}$ and $f_t = (f_{jt})_{j=1,q}$ respectively be the vector of observed variables $\{y_1, ..., y_p\}$ and latent factors $\{f_1, ..., f_q\}$ for observation t. We have:

$$\forall\, t\ f_t \sim \mathcal{N}(0, I_q)$$

Conditional to factors f_t, $(y_{it})_{i=1,p}$ are independently distributed according to a model having an exponential structure [12], i.e. the density of which has the form:

$$l_i(y_{it}|\delta_{it}, \phi) = \exp\left\{\frac{(y_{it}\delta_{it} - b_i(\delta_{it}))}{a_{it}(\phi)} + c_i(y_{it}, \phi)\right\} \tag{1}$$

From:

$$\begin{cases} \mathbb{E}\left(\frac{\partial l_i(y_{it}|\delta_{it},\phi)}{\partial \delta_i}\right) & = 0 \\ \mathbb{E}\left(\frac{\partial^2 l_i(y_{it}|\delta_{it},\phi)}{\partial \delta_i^2}\right) + \mathbb{E}\left(\left(\frac{\partial l_i(y_{it}|\delta_{it},\phi)}{\partial \delta_i}\right)^2\right) & = 0 \end{cases}$$

one gets the following useful classical results. $\forall i$, t

$$\mu_{it} = \mathbb{E}(y_{it}) = b_i'(\delta_{it}) \tag{2}$$

$$Var(y_{it}) = a_{it}(\phi) b_i''(\delta_{it}) = a_{it}(\phi) b_i'' \left[b_i'^{-1}(\mu_{it})\right] \text{ with } v_i(\mu_{it}) = b_i'' \left(\left[b_i^{-1\prime}(\mu_{it})\right]\right) \tag{3}$$

Independence of $(y_{it})_{i=1,q}$conditional to f_t implies that they have conditional variance matrix:

$$Var_{f_t}(y_t) = diag\left\{a_{it}(\phi) v_i(\mu_{it})\right\}_{i=1,...,q} \tag{4}$$

Table 1 gives the expressions of μ and $v(\mu)$ for the most usual distributions in the exponential family.

Table 1: μ and $v(\mu)$ for usual distributions in the exponential family

Distribution	μ	$a(\phi)$	$v(\mu)$
$a\mathcal{B}(n,p)$	$p = \frac{e^\delta}{1+e\delta}$	a	$\mu(1-\mu)$
$\mathcal{P}(\lambda)$	$\lambda = e^\delta$	1	μ
$\mathcal{G}\left(a, \frac{1}{\lambda}\right)$	$a\lambda = \frac{1}{\delta}$	1	$-\mu^2$
$\mathcal{N}(\mu, \sigma^2)$	$\mu = \delta$	σ^2	1

2.2 Linear predictors

For every $i = 1, p$, factors $\{f_1, ..., \ f_q\}$ generate a predictor η_i underlying variable y_i. This predictor is assumed to be a linear combination of the factors:

$$\forall\, i, t \quad \eta_{it} = \theta_i + a_i' f_t, \quad a_i \in \mathbb{R}^p \tag{5}$$

Let $\theta = (\theta_1, ..., \theta_p)'$ the vector of the predictor means. Most generally, these effects may depend on covariates, but in order to simplify our developments, we here take them constant. Let $F = (f_1, ..., f_t, ..., f_n)'$ be the (n, q) factor matrix and $A = (a_1, ..., a_p)'$ be the (p, q) factor coefficient matrix.

Let finally:

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

On observation level, (5) may then be written:

$$\forall\, t \quad \eta_t = \theta + A f_t \tag{6}$$

which, on global level, reads:

$$\underset{(p,n)}{\eta} = \underset{(p,1)}{\theta}\underset{(1,n)}{\mathbf{1}_n'} + \underset{(p,q)}{A}\underset{(q,n)}{F} \tag{7}$$

2.3 Link function

The linear predictor and the expectation of the dependent variable y_i are linked through a link function g_i:

$$\forall\, i, t \quad \eta_{it} = g_i(\mu_{it}) \tag{8}$$

Amongst all link functions, that which allows to equate the linear predictor η and the canonical parameter δ is called canonical link function. We have:

$$(2) \text{ and } (8) \ \Rightarrow \ \eta_{it} = g_i\left(b_i'(\delta_{it})\right)$$

So, the canonical link function is: $g_i = b_i'^{-1}$. Table 2 gives the canonical link functions of the most usual distributions in the exponential family.

3 Estimation of the GLFM

As, conditional to the factors, the GLFM boils down to a GLM, we first recall the overall structure of the GLM estimation algorithm, which also allows to introduce our notations. Then, we give back their latent random variable status to the factors, and adapt the estimation procedure to this situation by including an EM step in its current iteration.

Table 2: Canonical link functions of usual distributions in the exponential family

$a\mathcal{B}(n,p)$	$\mathcal{P}(\lambda)$	$\mathcal{G}(a,\frac{1}{\lambda})$	$\mathcal{N}(\mu,\sigma^2)$
$g(x)=\log\left(\frac{x}{1-x}\right)$	$g(x)=\log(x)$	$g(x)=\frac{1}{x}$	$g(x)=x$

3.1 Estimating a GLM through the Fisher's score algorithm

3.1.1 Univariate GLM

Consider the GLM of some variable y, with $\mu=\mathbb{E}(y)$. The explanatory variables $\{x_1,...,x_q\}$ are observed. Let:

$$\mathbf{x}_t=(x_{1t},...,x_{qt})'\ ;\ X=(x_1,...,x_q)=(\mathbf{x}_1,...,\mathbf{x}_t,...,\mathbf{x}_n)'$$

Let g be the link function, and η the linear predictor:

$$\eta=X\beta,\quad \beta\in\mathbb{R}^p$$

For each unit t, we have:

$$\eta_t=g(\mu_t)\quad\Rightarrow\quad \mathbf{x}_t'\beta=g\left(b'(\delta_t)\right)$$

The problem is to estimate β. The log-likelihood of the model is:

$$\mathcal{L}(\delta;y)=\sum_{t=1}^{n}\mathcal{L}_t(\delta_t;y_t)=\sum_{t=1}^{n}\left[\frac{y_t\delta_t-b(\delta_t)}{a_t(\phi)}+c(y_t,\phi)\right]$$

Derivation with respect to β yields:

$$\begin{aligned}\frac{\partial\mathcal{L}_t}{\partial\beta_j}&=\frac{\partial\eta_t}{\partial\beta_j}\frac{\partial\mu_t}{\partial\eta_t}\frac{\partial\delta_t}{\partial\mu_t}\frac{\partial\mathcal{L}_t}{\partial\delta_t}=x_{tj}\frac{1}{g'(\mu_t)}\frac{1}{b''(\delta_t)}\frac{y_t-\mu_t}{a_t(\phi)}\\ \Rightarrow\ \frac{\partial\mathcal{L}}{\partial\beta_j}&=\sum_{t=1}^{n}x_{tj}\frac{1}{g'(\mu_t)^2var(y_t)}g'(\mu_t)(y_t-\mu_t)\end{aligned}$$

Let:

$$W_\beta=diag\left[g'(\mu_t)^2var(y_t)\right]_{t=1,n}=diag\left[g'(\mu_t)^2a_t(\phi)v(\mu_t)\right]_{t=1,n}$$

The expression of W_β for usual models can be found in table 3.
Let also:

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

Table 3: Expression of W_β for usual models in the exponential family

Distribution	W_β
$a\mathcal{B}(n,p)$	$W_\beta = diag\left(a\frac{1+\exp(\mathbf{x}_t'\beta)}{\exp(\mathbf{x}_t'\beta)}\right)$
$\mathcal{P}(\lambda)$	$W_\beta = diag\left(\frac{1}{\exp(\mathbf{x}_t'\beta)}\right)$
$\mathcal{G}(a,\frac{1}{\lambda})$	$W_\beta = diag\left(a\left(\exp(\mathbf{x}_t'\beta)\right)^2\right)$
$\mathcal{N}(\mu,\sigma^2)$	$W_\beta = diag\left(\psi_t\right)$

Then, likelihood equations can be written:

$$\underset{\beta}{\nabla} L = 0 \;\Leftrightarrow\; X'W_\beta^{-1}\frac{\partial\eta}{\partial\mu}(y-\mu) = 0 \tag{9}$$

This equation system not being linear in β, it is solved using an iterative process, known as Fisher's scores algorithm. If $m^{[k]}$ denotes the value of element m after iteration k:

$$\begin{aligned}
\beta^{[k+1]} &= \beta^{[k]} - \left(\mathbb{E}\left[\left\{\frac{\partial^2\mathcal{L}}{\partial\beta\partial\beta'}\right\}\right]^{[k]}\right)^{-1}\left(\frac{\partial\mathcal{L}}{\partial\beta}\right)^{[k]} \\
&= \beta^{[k]} + \left(X'W_{\beta^{[k]}}^{-1}X\right)^{-1}X'W_{\beta^{[k]}}^{-1}\left(\frac{\partial\eta}{\partial\mu}\right)^{[k]}\left(y-\mu^{[k]}\right) \\
&= \left(X'W_{\beta^{[k]}}^{-1}X\right)^{-1}X'W_{\beta^{[k]}}^{-1}z^{[k]}
\end{aligned} \tag{10}$$

where:

$$z^{[k]} = X\beta^{[k]} + \left(\frac{\partial\eta}{\partial\mu}\right)^{[k]}\left(y-\mu^{[k]}\right)$$

Equation (10) may be interpreted as the normal equations of a linear model. Indeed, let:

$$z_\beta = \eta + \frac{\partial\eta}{\partial\mu}(y-\mu) = X\beta + \frac{\partial\eta}{\partial\mu}(y-\mu) \tag{11}$$

Then, (9) becomes:

$$X'W_\beta^{-1}(z_\beta - X\beta) = 0 \tag{12}$$

Equations (12) with given z_β may be interpreted as GLS normal equations of the following linear model:

$$\mathcal{M} \quad : \qquad z_\beta = X\beta + \zeta\,, \quad \text{where}: \ \mathbb{E}(\zeta) = 0 \ ; \ V(\zeta) = W_\beta$$
$$(\text{indeed: } V(\zeta_t) = V(z_{\beta,t}) = g'(\mu_t)^2 Var(y_t))$$

So, current iteration k of the estimation algorithm consists in solving $X'W^{-1}_{\beta^{[k]}}(z_{\beta^{[k]}} - X\beta) = 0$ with respect to β, and updating β in W_β and z_β with the solution.

We shall refer to $\mathcal{M}^{[k]} : z_{\beta^{[k]}} = X\beta + \zeta^{[k]}; \quad \mathbb{E}(\zeta^{[k]}) = 0; \quad V(\zeta^{[k]}) = W_{\beta^{[k]}}$ as the (current) *linearized model*. One important point is that GLS estimation of this model is nothing but a Quasi-Likelihood Estimation (QLE). This estimation by maximum of QL mimics MLE on each step, under a normality and independence assumption of the $z_{\beta,t}$'s with a fixed covariance structure.

Notes:

1. In the particular case of the normal distribution, the linearized model is no other than the initial linear model.
2. As the 1^{st} order development of g at point μ yields:

$$g(y) \approx g(\mu) + g'(\mu)(y - \mu) = z$$

we may perform OLSR of $g(y)$ on X, in order to get an initial value $\beta^{[0]}$. When $g(y)$ is not defined owing to zero-values in data, we have to mix y up with some relevant quantity. We propose to take:

$$\forall\, t = 1, n \quad z_t^{[0]} = g\left[\alpha y_t + (1-\alpha)\overline{y}\right], \quad \text{with } \alpha = 0.95$$

3.1.2 Multivariate GLM

Consider now that $y_1, ..., y_p$ depend on the same explanatory variables $\{x_1, \ ... \ , x_q\}$, conditional to which they are all independent.

The conditional independence assumption implies that:

$$\forall\, t = 1, n: \ l(y_t|\eta_t) = \prod_{i=1}^{p} l_i(y_{it}|\eta_{it})$$

That is, in view of the independence of units:

$$l(\mathcal{Y}|\eta) = \prod_{i=1}^{p} l_i(y_i|\eta_i)$$

As a result, the corresponding linearized model in the FSA is the following:

$$\mathcal{M}: \quad \forall\, i = 1, p: \quad z_{i\beta} = X\beta_i + \zeta_i$$

where the ζ_i's are independent and $\forall\, i$: $\mathbb{E}(\zeta_i) = 0$; $Var(\zeta_i) = W_{i\beta}$ with $W_{i\beta} = diag\left(g_i'(\mu_{it})^2 Var(y_{it})\right)_{t=1,n} = diag\left(g_i'(\mu_{it})^2 a_{it}(\phi) v(\mu_{it})\right)_{t=1,n}$.

The FSA is used to estimate this model.

3.2 Estimation of a normal FM

3.2.1 Homoskedastic model: basic application of the EM algorithm

Consider the following normal factor model:

$$\forall\, t = 1, n: \quad y_t = \theta + A f_t + \varepsilon_t\ ; \quad Var(\varepsilon_t | f_t) = \Psi$$

$$\begin{pmatrix} y_t \\ f_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \theta \\ 0 \end{pmatrix}; \begin{pmatrix} AA' + \Psi & A \\ A' & I_q \end{pmatrix} \right)$$

1^{st} **order equations:** A is estimated by using the expectation, conditional to observations, of the derivative with respect to parameters of the completed log-likelihood (EDLCO):

$$\sum_t \mathbb{E}\left(\underset{A,\Psi}{\nabla} \log l(y_t, f_t) | y_t \right) = 0$$

This can be achieved because EDLCO is analytically determined. Let us give back some landmarks in this EM estimation.

The completed log-likelihood of observations is:

$$-\frac{1}{2} \sum_{t=1}^{n} \left(\log|\Psi| + (y_t - A f_t - \theta)\Psi^{-1}(y_t - A f_t - \theta) \right) + R$$

where R is a term independent from θ and A. Setting to zero its derivative with respect to θ yields:

$$\sum_t (y_t - \widehat{A} f_t - \widehat{\theta}) = 0 \tag{13}$$

If f_t were known, (13) could be used in the following updating formula:

$$\widehat{\theta}^{[k+1]} = \frac{1}{n} \sum_t (y_t - \widehat{A}^{[k]} f_t^{[k]})$$

As it is not, EM replaces f_t with its expectation conditional to y, $\widetilde{f_t} = \mathbb{E}(f_t | y_t)$, which gives:

$$\widehat{\theta}^{[k+1]} = \frac{1}{n} \sum_t (y_t - \widehat{A}^{[k]} \widetilde{f}_t^{[k]}) \quad \text{where} \quad \widetilde{f}_t^{[k]} = \mathbb{E}(f_t^{[k]} | y_t) = \gamma(y_t - \widehat{\theta}^{[k]}) \tag{14}$$

In like manner, setting to zero the derivative with respect to A yields, and substituting $\widetilde{f_t}$ with gives:

$$\sum_t (y_t \widetilde{f}_t' - \widehat{\theta}\widetilde{f}_t' - \widehat{A}\widetilde{S}_t) = 0 \quad \text{where} \quad \widetilde{S}_t = \mathbb{E}(f_t f_t' | y_t) \tag{15}$$

(14) put into (15) yields:

$$\left[\sum_t y_t \widetilde{f}_t' - \frac{1}{n}\sum_t y_t \sum_t \widetilde{f}_t'\right] = \widehat{A}\left[\sum_t \widetilde{S}_t - \frac{1}{n}\sum_t \widetilde{f}_t \sum_t \widetilde{f}_t'\right]$$

$$\Leftrightarrow \widehat{A} = \left[\sum_t y_t \widetilde{f}_t' - \frac{1}{n}\sum_t y_t \sum_t \widetilde{f}_t'\right]\left[\sum_t \widetilde{S}_t - \frac{1}{n}\sum_t \widetilde{f}_t \sum_t \widetilde{f}_t'\right]^{-1} \tag{16}$$

with:

$$\widetilde{f}_t = \gamma(y_t - \widehat{\theta}) = \widehat{A}'\left(\widehat{A}\widehat{A}' + \widehat{\Psi}\right)^{-1}(y_t - \widehat{\theta}) \tag{17}$$

$$\widetilde{S}_t = I - \widehat{A}'\left(\widehat{A}\widehat{A}' + \widehat{\Psi}\right)^{-1}\widehat{A} + \widetilde{f}_t\widetilde{f}_t' \tag{18}$$

Derivation with respect to inverse of the variance matrix gives:

$$\underset{\Psi^{-1}}{\nabla} \sum_t \left[\log|\Psi| + \|y_t - \widehat{\theta} - \widehat{A}f_t\|^2_{\Psi^{-1}}\right] = 0$$

$$\Leftrightarrow \quad \widehat{\Psi} = \frac{1}{n}\sum_t (y_t - \widehat{\theta} - \widehat{A}f_t)(y_t - \widehat{\theta} - \widehat{A}f_t)'$$

the expectation of which is taken conditional to the observed data, giving:

$$\widehat{\Psi} = \frac{1}{n}\sum_t \left[(y_t - \widehat{\theta})(y_t - \widehat{\theta})' - \widehat{A}\widetilde{f}_t(y_t - \widehat{\theta})' - (y_t - \widehat{\theta})\widetilde{f}_t'\widehat{A} + \widehat{A}\widetilde{S}_t\widehat{A}'\right] \tag{19}$$

Estimation of A and Ψ: The solution of system (16-19) can be viewed as the fixed point of the following iterative procedure:

$$\widehat{A}^{[k+1]} = \left[\sum_t y_t \widetilde{f}_t^{[k]\prime} - \frac{1}{n}\sum_t y_t \sum_t \widetilde{f}_t^{[k]\prime}\right]\left[\widetilde{S}_t^{[k]} - \frac{1}{n}\sum_t \widetilde{f}_t^{[k]} \sum_t \widetilde{f}_t^{[k]\prime}\right]^{-1} \tag{20}$$

with:

$$\widetilde{f}_t^{[k]} = \gamma(y_t - \widehat{\theta}^{[k]}) = \widehat{A}^{[k]\prime}\left(\widehat{A}^{[k]}\widehat{A}^{[k]\prime} + \widehat{\Psi}^{[k]}\right)^{-1}(y_t - \widehat{\theta}^{[k]}) \tag{21}$$

$$\widetilde{S}_t^{[k]} = I - \widehat{A}^{[k]\prime}\left(\widehat{A}^{[k]}\widehat{A}^{[k]\prime} + \widehat{\Psi}^{[k]}\right)^{-1}\widehat{A}^{[k]} + \widetilde{f}_t^{[k]}\widetilde{f}_t^{[k]\prime} \tag{22}$$

$$\widehat{\Psi}^{[k+1]} = \frac{1}{n}\sum_t \left[(y_t - \widehat{\theta}^{[k]})(y_t - \widehat{\theta}^{[k]})' - \widehat{A}^{[k]}\widetilde{f}_t^{[k]}(y_t - \widehat{\theta}^{[k]})' - (y_t - \widehat{\theta}^{[k]})\widetilde{f}_t^{[k]\prime}\widehat{A}^{[k]} + \widehat{A}^{[k]}\widetilde{S}_t\widehat{A}^{[k]\prime}\right] \tag{23}$$

Estimation of F: f_t is estimated as its expectation conditional to y_t: $\widetilde{f}_t^{[\infty]}$.

3.2.2 Adaptation of the EM algorithm to the Heteroskedastic model

Following [3] and [15], consider now the normal FM:

$$\forall\, t = 1, n: \quad y_t = \theta + A f_t + \varepsilon_t\ ; \quad Var(\varepsilon_t | f_t) = \Psi_t \tag{24}$$

The completed log-likelihood of observations is:

$$-\frac{1}{2}\sum_{t=1}^{n}\left(\log|\Psi_t| + (y_t - A f_t - \theta)'\Psi_t^{-1}(y_t - A f_t - \theta)\right) + R \tag{25}$$

Setting to zero its derivative with respect to θ yields:

$$\sum_t \Psi_t^{-1}(y_t - \widehat{A} f_t - \widehat{\theta}) = 0 \tag{26}$$

So, we get:

$$\widehat{\theta} = \left(\sum_t \Psi_t^{-1}\right)^{-1} \sum_t \Psi_t^{-1}(y_t - \widehat{A} f_t) \tag{27}$$

leading to:

$$\widehat{\theta}^{[k+1]} = \left(\sum_t \Psi_t^{[k]-1}\right)^{-1} \sum_t \Psi_t^{[k]-1}(y_t - \widehat{A}^{[k]} \widetilde{f}_t^{[k]}) \tag{28}$$

In like manner, setting to zero the derivative with respect to A leads to:

$$\sum_t \Psi_t^{-1}\left[(y_t - \widehat{\theta})\widetilde{f}_t' - \widehat{A}\widetilde{S}_t\right] = 0 \tag{29}$$

$$\Leftrightarrow \sum_t \Psi_t^{-1}\left[(y_t - \widehat{\theta})\widetilde{f}_t'\right] = \sum_t \Psi_t^{-1}\widehat{A}\widetilde{S}_t \tag{30}$$

In case of a single factor, Ψ_t, $\widetilde{f}_t^{[k]}$, $\widetilde{S}_t^{[k]} \in \mathbb{R}$, so the solution of system (28-30) is straightforward:

(27) put into (29) yields:

$$\sum_t \Psi_t^{[k]-1}\left[\left(y_t - \left(\sum_t \Psi_t^{[k]-1}\right)^{-1}\sum_t \Psi_t^{[k]-1}(y_t - \widehat{A}^{[k]}\widetilde{f}_t^{[k]})\right)\widetilde{f}_t^{[k]\prime} - \widehat{A}^{[k]}\widetilde{S}_t^{[k]}\right] = 0 \tag{31}$$

which in this case, gives:

$$\begin{aligned}\widehat{A}^{[k]} &= \left[\sum_t \Psi_t^{[k]-1} y_t \widetilde{f}_t^{[k]\prime} - \frac{1}{M^{[k]}} \sum_t \Psi_t^{[k]-1} \left(\sum_t \Psi_t^{[k]-1} y_t\right) \widetilde{f}_t^{[k]\prime}\right] \\ &\times \left[\sum_t \Psi_t^{[k]-1} \widetilde{S}_t^{[k]} - \frac{1}{M^{[k]}} \sum_t \Psi_t^{[k]-1} \left(\sum_t \Psi_t^{[k]-1} \widetilde{f}_t^{[k]}\right) \widetilde{f}_t^{[k]\prime}\right]^{-1} \end{aligned} \tag{32}$$

where $M^{[k]} = \sum_t \Psi_t^{[k]-1}$.

Let us now consider the case of several factors (we shall take $k = 2$ to make writing simpler, but the technique holds in the most general situation). We can reason row by row. In view of the fact that $\Psi_t^{-1} = diag\left(\Psi_t^i\right)_i$ (independence of $\{y_1, ..., y_p\}$ conditional to $\{f_1, ..., f_q\}$) , row i of equation (30) boils down to:

$$\sum_t \frac{y_{it} - \theta_i}{\Psi_t^i} \begin{pmatrix} \widetilde{f}_t^1 & \widetilde{f}_t^2 \end{pmatrix} = \begin{pmatrix} \widehat{a}_{i1} & \widehat{a}_{i2} \end{pmatrix} \sum_t \frac{1}{\Psi_t^i} \begin{pmatrix} \widetilde{S}_t^{11} & \widetilde{S}_t^{12} \\ \widetilde{S}_t^{21} & \widetilde{S}_t^{22} \end{pmatrix}$$

where $\widehat{A} = \left(\left(\widehat{a}_{ij}\right)\right)_{i,j}$ and $\widetilde{S}_t = \left(\left(\widetilde{S}_t^{ij}\right)\right)_{i,j}$

We get the corresponding iterative equation system:

$$\sum_t \frac{y_{it} - \theta_i^{[k]}}{\Psi_t^{i[k]}} \begin{pmatrix} \widetilde{f}_t^{1[k]} & \widetilde{f}_t^{2[k]} \end{pmatrix} = \begin{pmatrix} \widehat{a}_{i1}^{[k+1]} & \widehat{a}_{i2}^{[k+1]} \end{pmatrix} \sum_t \frac{1}{\Psi_t^{i[k]}} \begin{pmatrix} \widetilde{S}_t^{11[k]} & \widetilde{S}_t^{12[k]} \\ \widetilde{S}_t^{21[k]} & \widetilde{S}_t^{22[k]} \end{pmatrix}$$

Solving this system yields row i of $\widehat{A}^{[k+1]}$.

■ Derivation with respect to inverse of the variance matrix gives:

$$\underset{\Psi_t^{-1}}{\nabla} \left[\log|\Psi_t| + \|y_t - \widehat{\theta} - \widehat{A} f_t\|^2_{\Psi_t^{-1}}\right] = 0$$

$$\Leftrightarrow \widehat{\Psi}_t = (y_t - \widehat{\theta} - \widehat{A} f_t)(y_t - \widehat{\theta} - \widehat{A} f_t)'$$

the expectation of which is taken conditional to the observed data, giving:

$$\widehat{\Psi}_t = \left[(y_t - \widehat{\theta})(y_t - \widehat{\theta})' - \widehat{A}\widetilde{f}_t(y_t - \widehat{\theta})' - (y_t - \widehat{\theta})\widetilde{f}_t'\widehat{A} + \widehat{A}\widetilde{S}_t\widehat{A}'\right] \tag{33}$$

Of course, if Ψ_t can be expressed as a function of other parameters, e.g. μ_t, such an expression may provide a better estimator than that of (33).

3.2.3 Identification constraints

As specified above (§3.2.1), the model is not identifiable: factors are defined except for an orthogonal transform P. Indeed:

$$Af_t = A^* f_t^* \quad \text{with} \quad A^* = AP^{-1} \quad \text{and} \quad f_t^* = Pf_t$$

And

$$Var(f_t^*) = PVar(f_t)P' = PP' = I_q$$

So, models:

$$y_t = \theta + Af_t + \varepsilon_t \quad \text{and} \quad y_t = \theta + A^* f_t^* + \varepsilon_t \quad \text{with} \quad Var(\varepsilon | f_t) = \Psi$$

meet exactly the same hypotheses. To make factors identifiable, we must impose constraints to coefficient matrix A. There are several options that may be taken [refs]. We used that of hierarchical constraints, explained hereafter.

Matrix A has dimensions (p, q) with $q \leq p$. So, we can write, in a unique manner:

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad \text{where } A_1 \text{ is a } (q, q) \text{ matrix}$$

Now, assuming that A_1 is a full rank matrix, there is a unique orthogonal transform P such that $A_1 P'$ is a lower triangular matrix with positive diagonal elements.

Proof: $M = A_1 A_1'$ is symmetric definite positive, so it can be written:

$$M = LDL'$$

where L is lower triangular and D is diagonal with positive diagonal elements. Then $L_1 = LD^{\frac{1}{2}}$ is the unique lower triangular matrix with positive diagonal elements that verifies Cholesky's decomposition:

$$M = L_1 L_1'$$

So, $P = L_1^{-1} A_1$ is such that:

$$PP' = L_1^{-1} A_1 A_1' L_1^{-1\prime} = L_1^{-1} LDL' L_1^{-1\prime} = D^{-\frac{1}{2}} L^{-1} LDL' L^{-1\prime} D^{-\frac{1}{2}} = I_q$$

and

$$A_1 P' = A_1 P^{-1} = A_1 A_1^{-1} L_1 = L_1$$

So, in order to make the model identifiable, we constrain coefficient matrix A to have the following form:

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad \text{where } A_1 \text{ is a lower triangular matrix}$$

$$\Leftrightarrow \quad A = ((a_{ij}))_{\substack{i=1,p \\ j=1,q}} \quad \text{with } a_{ij} = 0 \; \forall \, j > i$$

3.2.4 Constrained EM estimation

Following [1] and [15], the hierarchical constraint proves handy: while solving (30) row by row, one just has to set to zero all coefficients $\widehat{a}_{ij}$ that must equal zero in A_1. More precisely, if in row i, coefficients $\{a_{ij}, j > i\}$ are to be 0, then only the first i elements of row i are kept, both in matrices $\sum_t \Psi_t^{-1} \widehat{A} \widetilde{S}_t$ and $\sum_t \Psi_t^{-1} \left[(y_t - \widehat{\theta}) \widetilde{f}_t' \right]$, providing the equation system yielding coefficients $\{\widehat{a}_{ij}, j \leq i\}$.

3.3 Estimation of a GLFM

3.3.1 The principle

So, in the case of a classical FM, the EM algorithm takes advantage of the fact that all variables are normally distributed to use analytic expressions of all required expectations conditional to y. According to section 2, such a normality assumption may be formally used with the linearized GLM within current step k, since GLS mimics normal MLE.

In the case of a GLFM, the estimation principle we propose is then informally straightforward. We consider the model alternately as:

- a GLM model conditional to F in the *linearization* step.
- a FM within the current *estimation step* of this GLM, as this step uses the previously produced linearized version of the GLM.

3.3.2 The linearization step

Conditional to the current values of $\theta\,, A\,, F$ and following (11), we introduce the pseudo-dependent working variable z, which is then known:

$$\forall \, i = 1, p: \quad z_{i,F} = \theta_i \mathbf{1}_n' + F a_i + \frac{\partial \eta_{i,F}}{\partial \mu_{i,F}} (y_i - \mu_{i,F}) = \theta_i \mathbf{1}_n' + F a_i + g'(\mu_{i,F})(y_i - \mu_{i,F})$$

$$\text{let} \quad \zeta_{i,F} = g'(\mu_{i,F})\epsilon_{i,F} \quad \text{with} \quad \epsilon_{i,F} = y_i - \mu_{i,F}$$

Let $\forall\, t,\ z_t = (z_{1t}, ..., z_{pt})'$, and $Z = (z_1, ..., z_t, ..., z_n)'$. Given Z, F and $Var(\zeta|F)$, we have the *linearized conditional model*:

$$\forall\, t = 1, n: \quad z_t = \theta + Af_t + \zeta_t \tag{34}$$

Of course, according to section 3.1.1, the variance of residuals in this model is such that:

$$Var(\zeta|F) = diag\left(g'(\mu_{it})^2 Var(\epsilon_{it|F})\right)_{t=1,n} \tag{35}$$

In the case of the canonical link, we have: $Var(\zeta_{iF|F}) = W_{iAF}$ and expressions of W corresponding to the usual distributions can again be found in table 3.

3.3.3 The estimation step

Given Z and $Var(\zeta)$, we now give back its random nature to f_t , and thus, view (34) as a FM called the *linearized marginal model* [7]. Its expectation is:

$$\forall\, t: \quad f_t \sim \mathcal{N}(0, I_k) \quad \Rightarrow \quad \mathbb{E}(Z_t) = \theta \tag{36}$$

But this is a heteroskedastic model, since its variance structure depends on the observation. We shall, in this step, consider it known, its value being estimated by (35). So, we shall take:

$$\Psi_t = g'^2(\mu_{i,f_t})Var(\epsilon_{it}|f_t)$$

If g is the canonical link function, we have:

$$Var(\epsilon_{it}|f_t) = Var(y_{it}|f_t) = a_{it}(\phi)b_i''\left(b_i'^{-1}(\mu_{it})\right) = a_{it}(\phi)g_i^{-1\prime}\left(g_i(\mu_{it})\right) = a_{it}(\phi)/g_i'(\mu_{it})$$

$$\Rightarrow \quad \Psi_t = a_{it}(\phi)g'(\mu_{i,f_t})$$

3.3.4 The algorithm

To sum things up, here is the structure of the algorithm:

(0) Initialization:

Calculate:

$$\forall\, t = 1, n;\ \forall\, i = 1, p: \quad z_{it}^{[0]} = g(\alpha y_{it} + (1-\alpha)\overline{y}_i), \quad \text{with} \quad \alpha = 0.95$$

and, for instance:

$$\forall\, t = 1, n: \quad \Psi_t^{[0]} = I_p$$

N.B. If g is the canonical link, $\Psi_t^{[0]}$ may be initialized more accurately with:

$$\forall\, t = 1, n: \quad \Psi_t^{[0]} = diag\left(g_i'(m_{it})^2 a_{it}(\phi)\nu(m_{it})\right)_{i=1,p}$$

$$\text{where} \quad m_{it} = \alpha y_{it} + (1-\alpha)\overline{y}_i$$

(i) Given Z and $Var(\zeta)$, we have the linearized marginal model:

$$\forall\, t = 1, n: \quad z_t = \theta + A f_t + \zeta_t$$

viewed as a non-standard FM (with heterogenous variance structure $Var(\zeta_t) = \Psi_t$), estimated through an EM step, yielding F.

(ii) Given F, we have the linearized *conditional* model, viewed as a GLM: FSA updates θ and A using variance matrix:

$$Var(z_t|F_t) = Var(\zeta_t) = \Psi_t$$

(iii) Conditional to θ, A, F, calculate Z and $Var(\zeta)$:

$$\begin{array}{l} \forall\, i = 1, p: \quad \epsilon_{i,F} = y_i - \mu_{i,F} \quad ; \quad \zeta_{i,F} = g'(\mu_{i,F})\epsilon_{i,F} \quad ; \quad z_{i,F} = \theta_i \mathbf{1}_n + F a_i + \zeta_{i,F} \\ \forall\, t = 1, n: \quad Var(\zeta_t) = \Psi_t = diag\left(a_{it}(\phi) g'(\mu_{i,f_t})\right)_{i=1,p} \end{array}$$

Go to (i).

4 Experimental results

We present simulations carried out on a GLFM with one, two and three common latent factors, based respectively on the Poisson and the binomial distributions ($g = \log x$, respectively $\log \frac{x}{1-x}$). The simulated data vector has size $q = 40$ with $n = 800$. The convergence threshold N was taken equal to 10^{-5}.[1]

$$N = \max_{i \in \{1,\ldots,k\}} \left\{ \sum_{t=1}^{n} \left(f_{it}^{[e+1]} - f_{it}^{[e]} \right)^2 \right\}$$

Initial parameter values for the EM algorithm were obtained through random perturbation of the real parameter values.

4.1 Example 1: a Poisson GLFM

As EM also requires an initial value for z, we used the following approximation:

$$\forall\, i = 1, q\,; t = 1, n \quad z_{it}^{[0]} = \log\left[\alpha y_{it} + (1-\alpha)\overline{y}_i\right] , \quad \text{with } \alpha = 0.95$$

The rationale behind the use of $\alpha < 1$ is to circle difficulties due to zero-values in data. Our tests showed a good behaviour of the algorithm both at parameter and factor estimation.

Results from the regression of the simulated factors f_t on the estimated factors $\widetilde{f}_t$ (e.g., $f_{1t} = \beta_1 \widetilde{f}_{1t} + \gamma_1 \widetilde{f}_{2t} + \delta_1 \widetilde{f}_{3t} + \nu_t$, $f_{2t} = \beta_2 \widetilde{f}_{1t} + \gamma_2 \widetilde{f}_{2t} + \delta_2 \widetilde{f}_{3t} + \nu_t$ and $f_{3t} = \beta_3 \widetilde{f}_{1t} + \gamma_3 \widetilde{f}_{2t} + \delta_3 \widetilde{f}_{3t} + \nu_t$) given in figure 1 show that the regression coefficients β_1, γ_2 and δ_3 converge to one, while γ_1, δ_1, β_2, δ_2, β_3 and γ_3 are close to zero. This figure shows also that the correlations between simulated factors and their estimation was very close to 1 ($r_{f_1,\widetilde{f}_1}$, $r_{f_2,\widetilde{f}_2}$, $r_{f_3,\widetilde{f}_3} > 90\%$, and $r_{f_1,\widetilde{f}_2} \approx r_{f_2,\widetilde{f}_1} \approx \ldots \approx 0$). In this case the convergence threshold was reached after approximately 35 iterations (Figure 3).

4.2 Example 2: a Binomial GLFM

All the results are given in figures 2, 3 and 4 and tables 6, 7 and 8.

Results from the regression of the simulated factors f_t on the estimated factors $\widetilde{f}_t$ given in figure 2 show that the regression coefficients β_1, γ_2 and δ_3 converge to one, while γ_1, δ_1, β_2, δ_2, β_3 and γ_3 are close to zero. This figure shows also that the correlations between simulated factors and their estimation was very close to 1. Figure 3 shows that the convergence threshold was reached after approximately 28 iterations.

[1] $[e]$ is the iteration number

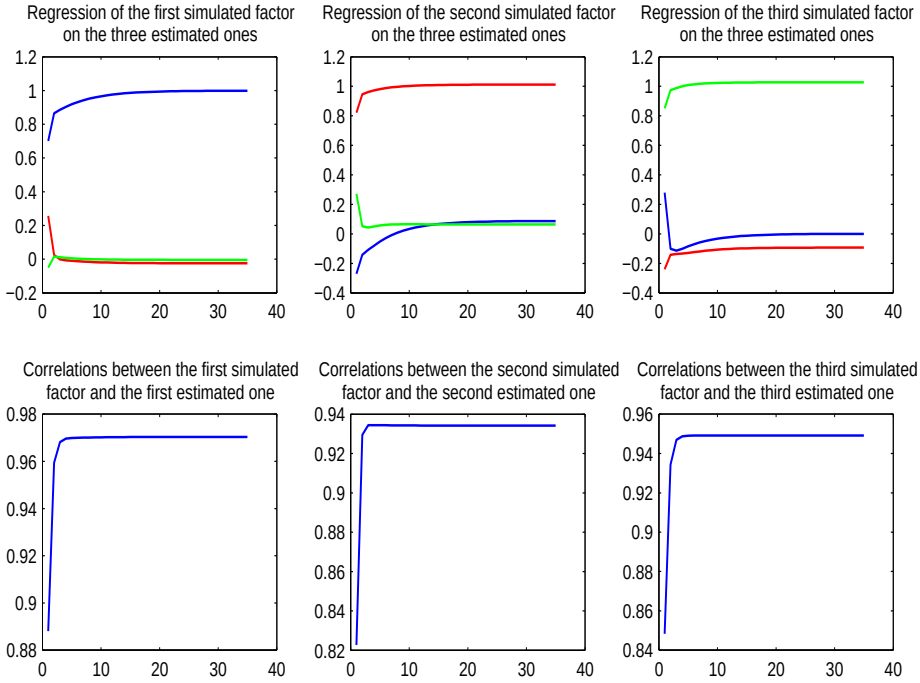

Figure 1: (a Poisson GLFM $\mathcal{P}(\lambda)$ with $k = 3$ common latent factors). In the first panel, blue lines represent the behaviour of the regression coefficient of the first factor; red lines the second factor and green lines the third one. The second panel shows the behaviour of the correlations coefficients.

Table 4: The correlations between the estimated and true parameter values in the case of a Poisson GLFM $\mathcal{P}(\lambda)$ with $k = 3$ common latent factors.

Vrai modèle	Modèle estimé	m	X		
1F	1F	0.9902	0.9939		
	2F	0.9898	0.9934		
	3F	0.9873	0.9866		
2F	1F	0.9696	0.6857		
	2F	0.9830	0.9884	0.9954	
	3F	0.9829	0.9880	0.9904	
3F	1F	0.9066	0.6735		
	2F	0.9536	0.9776	0.8373	
	3F	0.9898	0.9984	0.9936	0.9978

5 Conclusion

This paper discusses generalized linear latent factor models GLFMs as a tool to model longitudinal and (other forms of) clustered data. In sections 2 and 3 the most important concepts on model formulation, estimation, inference and prediction are summarized. A

Table 5: The parameter estimation errors in the case of a Poisson GLFM $\mathcal{P}(\lambda)$ with $k = 3$ common latent factors.

Vrai modèle	Modèle estimé	m	X
1F	1F	0.0038	0.0042
	2F	0.0040	0.0043
	3F	0.0048	0.0079
2F	1F	0.0073	0.3238
	2F	0.0068	0.0095
	3F	0.0078	0.0115
3F	1F	0.0823	1.8412
	2F	0.0255	1.2914
	3F	0.0085	0.3240

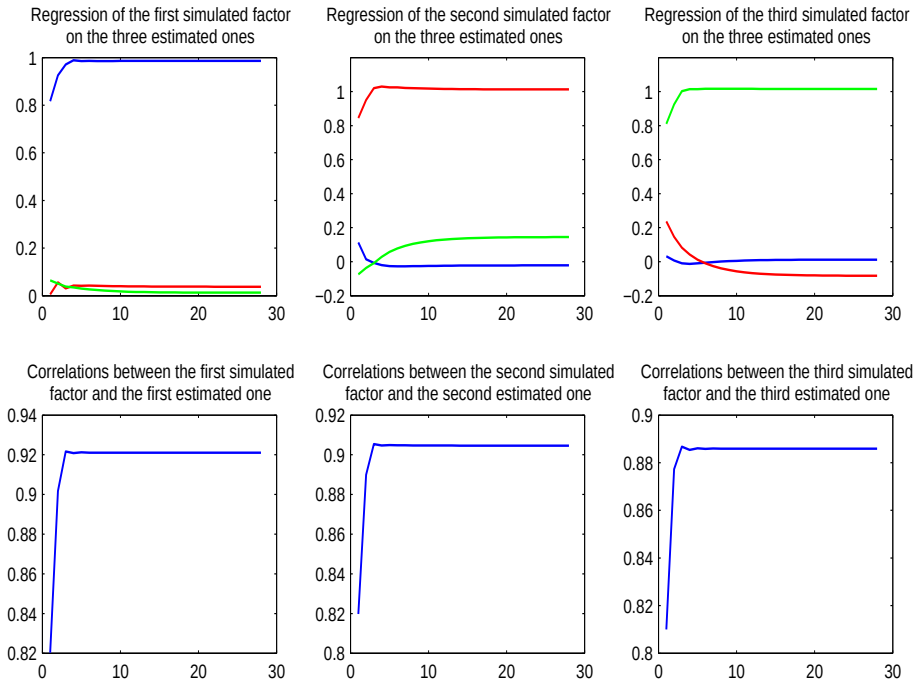

Figure 2: (a Binomial GLFM $\mathcal{B}(3, p)$ with $k = 3$ common latent factors). In the first panel, blue lines represent the behaviour of the regression coefficient of the first factor; red lines the second factor and green lines the third one. The second panel shows the behaviour of the correlations coefficients.

new estimation method combining the Fisher's score algorithm and a local EM inference step is presented.

Our proposed algorithm has been tested on simulated data and it showed very promising results. There are several benefits to use a local EM approximation for GLFM's. Most of these advantages revolve around the tractability of the learning and inference pro-

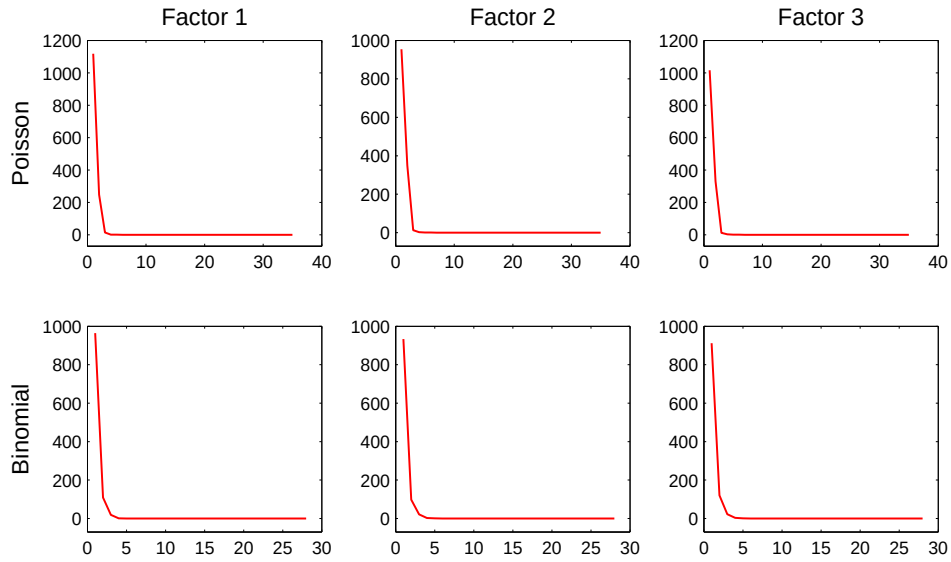

Figure 3: Behaviour of the convergence threshold N.

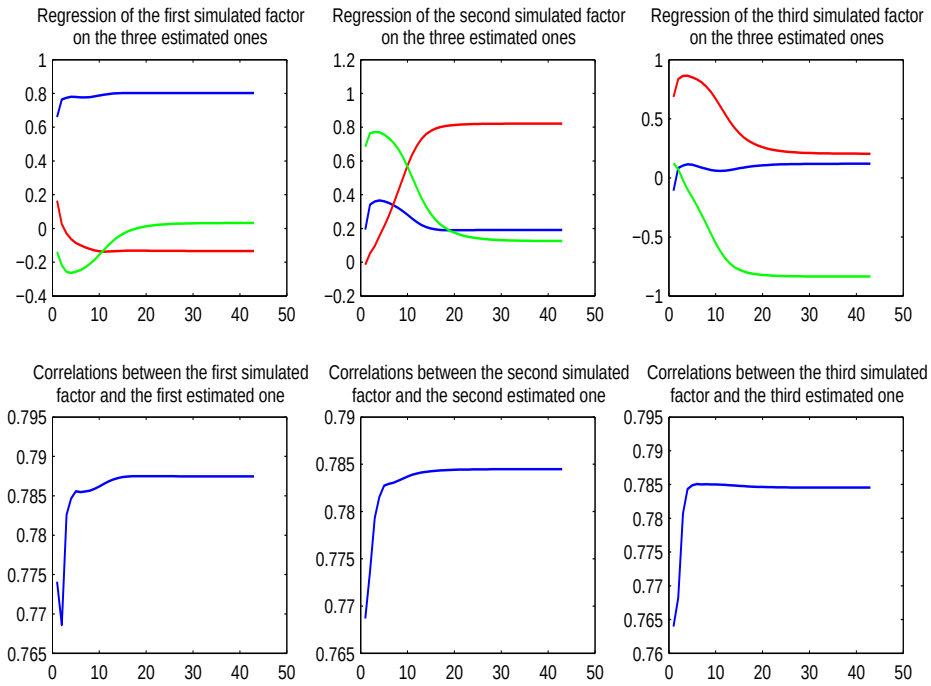

Figure 4: (Estimation of a Poisson GLFM $\mathcal{P}(\lambda)$ with $k = 3$ common latent factors using simulated data from a Binomial GLFM $\mathcal{B}(3, p)$ with $k = 3$). In the first panel, blue lines represent the behaviour of the regression coefficient of the first factor; red lines the second factor and green lines the third one. The second panel shows the behaviour of the correlations coefficients.

Table 6: The correlations between the estimated and true parameter values in the case of a Binomial GLFM $\mathcal{B}(3,p)$ with $k=3$ common latent factors.

Vrai modèle	Modèle estimé	m	X		
1F	1F	0.9776	0.9971		
	2F	0.9797	0.9968		
	3F	0.9804	0.9963		
2F	1F	0.6771	0.6572		
	2F	0.9406	0.9737	0.9966	
	3F	0.9368	0.9734	0.9955	
3F	1F	0.8203	0.4541		
	2F	0.8636	0.2012	0.8287	
	3F	0.9813	0.9856	0.9830	0.9902

Table 7: The parameter estimation errors in the case of a Binomial GLFM $\mathcal{B}(3,p)$ with $k=3$ common latent factors.

Vrai modèle	Modèle estimé	m	X
1F	1F	0.0048	0.0066
	2F	0.0055	0.0080
	3F	0.0084	0.0105
2F	1F	0.0945	0.2575
	2F	0.0427	0.0251
	3F	0.0567	0.0281
3F	1F	0.0982	0.1420
	2F	0.0494	0.4542
	3F	0.0225	0.0331

Table 8: The estimation errors and the correlations between the estimated and true parameter values in the case of the estimation of a Poisson GLFM $\mathcal{P}(\lambda)$ with $k=3$ common latent factors using simulated data from a Binomial GLFM $\mathcal{B}(3,p)$ with $k=3$. Values into brackets represent the results from the true specification.

	m	X		
Estimation errors	1.7023	0.7084		
	(0.0225)	(0.0331)		
Correlations	0.7020	0.8274	0.8901	-0.9496
	(0.9813)	(0.9856)	(0.9830)	(0.9902)

cesses and the increase in the processing speed and the improvement in the convergence performance.

Further empirical work should apply these models to the study of other financial and actuarial data. The main merits of GLFM's in this context are twofold. Firstly, regression

is no longer restricted to normal data, but extended to distributions from the exponential family. This enables appropriate modelling of, for instance, frequency counts, skewed or binary data. Secondly, a GLFM models the additive effect of explanatory variables and common latent factors on a transformation of the mean, instead of the mean itself. In addition, the approach adopted in this paper can be extended to allow for time-varying and more than one latent state variables as well as other types of probability distributions for the state variables. These extensions leave several interesting and challenging areas for future research.

References

[1] Aguilar, O. et West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics* **18**, 338–357.

[2] Bartholomew, D. J., and Knott, M., *Latent Variable Models and Factor Analysis*, London: Arnold (1999).

[3] Demos A., and Sentana E. (1998). An EM Algorithm for Conditionally Heteroscedastic Factor Models. *Journal of Business & Economic Statistics* **16** (3), 357–361.

[4] Gallant, A. R., and Tauchen, G. (1996). Which Moments to Match. *Econometric Theory*, **12**, 657–681.

[5] Genton, M.G., and Ronchetti, E. (2003). Robust Indirect Inference. *Journal of the American Statistical Association*, **98**, 1–10.

[6] Gouriéroux, C., Monfort, A., and Renault, A.E. (1993). Indirect Inference. *Journal of Applied Econometrics*, **8** (suppl.), S85–S118.

[7] Lavergne C. and Trottier C. (2000). Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires. *Revue de Statistique Appliquée*, **XLVIII**, 49–67.

[8] McLachlan, G.J., and Krishnan, T. (2008). *The EM algorithm and extensions*. Wiley Series in Probability and Statistics. New York, NY: John Wiley & Sons, Inc.

[9] Moustaki, I., "A Latent Trait and a Latent Class Model for Mixed Observed Variables", *British Journal of Mathematical and Statistical Psychology*, 49, 313–334 (1996).

[10] Moustaki, I., and Knott, M., "Generalized Latent Trait Models", *Psychometrika*, 65, 391–411 (2000).

[11] Moustaki, I and Victoria-Feser, M.P., "Bounded-influence robust estimation in generalized linear latent variable models", *Journal of the American Statistical Association*, 104, 644–653 (2006).

[12] Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A* **135**, 370–384.

[13] Rubin, D.B., and Thayer, D.T. (1982). EM algorithms for ML factor analysis. *Psychometrika* **47**, 69–76.

[14] Rubin, D.B., and Thayer, D.T. (1983). More on EM for ML factor analysis. *Psychometrika* **48**, 253–257.

[15] Saidane, M. and Lavergne, C. (2009). *"Modelling and Forecasting Volatility Dynamics Using Quadratic GARCH-Factor Models: Empirical Evidence from International Foreign Exchange Markets"* in Stock Returns: Cyclicity, Prediction and Economic Consequences (Editor: George I. Ellison), Nova Science Publishers, Inc. ISBN: 978-1-60741-458-2, Chapter 12, pp. 231–268 .

[16] Sammel, M. D., Ryan, L.M., and Legler, J.M., "Latent Variable Models for Mixed Discrete and Continuous Outcomes", *Journal of the Royal Statistical Society, Ser. B*, 59, 667–678 (1997).

[17] Skrondal, A., and Rabe-Hesketh, S., *Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models*, London: Chapman & Hall (2006).

[18] Wedel, M. and Kamakura, W. (2001). Factor Analysis with Mixed Observed and Latent Variables in the Exponential Family. *Psychometrika* **66** 515–30.