Bernoulli **20**(3), 2014, 1507–1531
DOI: 10.3150/13-BEJ531

Asymptotic properties of adaptive maximum likelihood estimators in latent variable models

SILVIA BIANCONCINI

Department of Statistical Sciences, University of Bologna, Via delle Belle Arti, 41-40126 Bologna, Italy.
E-mail: silvia.bianconcini@unibo.it

Latent variable models have been widely applied in different fields of research in which the constructs of interest are not directly observable, so that one or more latent variables are required to reduce the complexity of the data. In these cases, problems related to the integration of the likelihood function of the model arise since analytical solutions do not exist. In the recent literature, a numerical technique that has been extensively applied to estimate latent variable models is the adaptive Gauss–Hermite quadrature. It provides a good approximation of the integral, and it is more feasible than classical numerical techniques in presence of many latent variables and/or random effects. In this paper, we formally investigate the properties of maximum likelihood estimators based on adaptive quadratures used to perform inference in generalized linear latent variable models.

Keywords: Gaussian quadrature; generalized linear models; Laplace approximation; M-estimators

1. Introduction

Models based on latent variables are used in many scientific fields, particularly in social sciences. For instance, in psychology, researchers often use concepts as intelligence and anxiety, that are difficult to observe directly, but that can be indirectly measured by surrogate data based on individual responses to a battery of tests. In economics, welfare and poverty cannot be measured directly; hence income, expenditure and various other indicators on households are used as substitutes. Factor analysis is probably the best known latent variable model, based on the assumption of multivariate normality for the distribution of the manifest and latent variables. It has been extended by numerous researchers in order to deal with survey data that generally contain variables measured on binary, categorical or metric scales, or combinations of the above. Moustaki and Knott [11] proposed a Generalized Linear Latent Variable Model (GLLVM) framework that allows the distribution of the manifest variables to belong to the exponential family, that is either continuous or discrete variables.

The purpose of GLLVM is to describe the relationship between a set of responses or items $y_1, \dots, y_p$, and a set of latent variables or factors $z_1, \dots, z_q$, that are fewer in number than the observed variables. The factors are supposed to account for the dependencies among the response variables in the sense that if the factors are held fixed, then the observed variables are independent. This is known as the assumption of conditional or local independence. The conditional distribution of $y_j|\mathbf{z}$ ($\mathbf{z} = [z_1, \dots, z_q]^T$) is taken from the exponential family (with canonical link

1350-7265

functions)

$$g_j(y_j|\mathbf{z}) = \exp\left\{\frac{y_j(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z}) - b_j(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z})}{\phi_j} + c_j(y_j, \phi_j)\right\}, \qquad j = 1, \ldots, p,$$

where α_{0j} is the item-specific intercept, $\boldsymbol{\alpha}_j = [\alpha_{j1}, \ldots, \alpha_{jq}]^T$ can be interpreted as factor loadings of the model, and ϕ_j is the scale parameter, that is of interest in the case of continuous observed components. The functions $b_j(\cdot)$ and $c_j(\cdot, \cdot)$ are known and assume different forms according to the different nature of y_j.

Under the assumption of conditional independence, the joint marginal distribution of the manifest variables is

$$f(\mathbf{y}; \boldsymbol{\theta}) = \int_{\mathbb{R}^q} g(\mathbf{y}|\mathbf{z}; \boldsymbol{\theta}) h(\mathbf{z})\, \mathrm{d}\mathbf{z} = \int_{\mathbb{R}^q} \left[\prod_{j=1}^{p} g_j(y_j|\mathbf{z}; \boldsymbol{\theta})\right] h(\mathbf{z})\, \mathrm{d}\mathbf{z} \tag{1.1}$$

with $\mathbf{y} = [y_1, \ldots, y_p]^T$, $\boldsymbol{\theta} = [\alpha_{01}, \ldots, \alpha_{0p}, \boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_p^T, \phi_1, \ldots, \phi_p]^T$, and where $\mathbf{z}$ is generally assumed to be multivariate standard normal, but the independence assumption of the latent variables could be relaxed.

GLLVMs are designed as a flexible modelling approach. As a consequence, they are rather complex models, and their statistical analysis presents some difficulties due to the fact that the latent variables are not observed. Maximum likelihood estimates in the GLLVM framework are typically obtained by using standard maximization algorithms, such as the EM and the Newton–Raphson algorithms (Moustaki and Knott [11], Huber, Ronchetti and Victoria-Feser [5]). In both cases, the latent variables must be integrated out from the likelihood function, and numerical techniques have to be applied. Moustaki and Knott [11] proposed the use of the Gauss–Hermite (GH) quadrature as a numerical approximation method. Although this is feasible in fairly simple models and tends to work well with moderate sample sizes, its application is often unfeasible when the number of latent variables increases. Moreover, GH can completely miss the maximum for certain functions and can be inefficient in other cases. To overcome these limitations, the Adaptive Gauss–Hermite (AGH) quadrature has become very popular in the latent variable literature. It allows to get a better approximation of the integral by adjusting the quadrature locations with specific features of the posterior density of the latent variables given the observations. Developed in the Bayesian context by Naylor and Smith [13], it has been extended by several authors to deal with generalized linear mixed models. In particular, Schilling and Bock [23] applied the AGH quadrature to approximate marginal likelihoods in IRT models with binary data, whereas Rabe-Hesketh, Skrondal and Pickles [18] analyzed its behavior for generalized linear latent and mixed models. Furthermore, Joe [7] compared the AGH with the Laplace approximation for a variety of discrete response mixed models. He found that the Laplace approximation becomes less adequate as the degree of discreteness increases and suggests using AGH with binary and ordinal data. On this regard, we recall that several approaches have been proposed to overcome the main limitations of the Laplace approximation. In latent Gaussian models (Rue and Held [21]), the Integrated Nested Laplace Approximation (INLA) has become very popular to perform Bayesian inference with non-Gaussian observations (Rue, Martino and Chopin [22]). This procedure combines Laplace approximations with numerical integration to provide a fast and accurate method for approximating the predictive density of the latent variables/random effects. It is also a valuable tool in practice via the R-package *R-INLA* (Martins *et al.* [10]).

The adaptive Gauss–Hermite quadrature is implemented in many statistical software used to fit GLLVM, such as in the function *gllamm* in STATA (Rabe-Hesketh and Skrondal [17]), in MPLUS (Muthen and Muthen [12]), and in the PROC NLMIXED in SAS (Lesaffre and Spiessens [8]). However, to the best of our knowledge, inferential issues on the properties of the estimators based on the adaptive quadrature have not been addressed in the literature. In this paper, we formally investigate these theoretical properties as function of both the sample size and the number of observed variables. Our results generalize those by Huber, Ronchetti and Victoria-Feser [5], who analyzed the properties of classical Laplace-based estimators in GLLVM. Indeed, we show that the adaptive Gauss–Hermite quadratures share the same error rate of the higher (than one) order Laplace approximation.

The paper is organized as follows. In Section 2, we discuss the estimation of GLLVMs when the adaptive Gauss–Hermite quadrature is applied to approximate integrals. In Section 3, the relationship between AGH quadratures and the Laplace approximation is analyzed, and the asymptotic properties of the adaptive Maximum Likelihood (ML) estimators are derived. A simulation study is implemented in Section 4 to analyze the finite sample properties of the estimators. Finally, in Section 5, a brief summary on the main findings of the paper is provided.

2. Estimation based on adaptive Gauss–Hermite quadrature

Maximum Likelihood (ML) estimates in the GLLVM framework are typically obtained by using either the EM or the Newton–Raphson algorithms. The key component for applying both the algorithms is the score vector of the observed data log-likelihood function. For a random sample of size n, the latter is defined as

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \sum_{l=1}^{n} \log f(\mathbf{y}_l;\boldsymbol{\theta}) \\ &= \sum_{l=1}^{n} \log \int_{\mathbb{R}^q} \prod_{j=1}^{p} \exp\left[\frac{y_{jl}(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}_l)-b_j(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}_l)}{\phi_j}+c_j(y_{jl},\phi_j)\right] \\ &\qquad \times (2\pi)^{-q/2}\exp\left[-\frac{1}{2}\mathbf{z}_l^T\mathbf{z}_l\right]\mathrm{d}\mathbf{z}_l. \end{aligned} \tag{2.1}$$

It is easily shown that the score vector corresponding to expression (2.1) equals

$$\begin{aligned} S(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \sum_{l=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\log f(\mathbf{y}_l;\boldsymbol{\theta}) \\ &= \sum_{l=1}^{n}\frac{1}{f(\mathbf{y}_l;\boldsymbol{\theta})}\int_{\mathbb{R}^q}\frac{\partial}{\partial\boldsymbol{\theta}}\big[g(\mathbf{y}_l|\mathbf{z}_l;\boldsymbol{\theta})h(\mathbf{z}_l)\big]\,\mathrm{d}\mathbf{z}_l \end{aligned} \tag{2.2}$$

$$= \sum_{l=1}^{n} \frac{\int_{\mathbb{R}^q} S_l(\boldsymbol{\theta}; \mathbf{z}_l) g(\mathbf{y}_l|\mathbf{z}_l; \boldsymbol{\theta}) h(\mathbf{z}_l)\, \mathrm{d}\mathbf{z}_l}{\int_{\mathbb{R}^q} g(\mathbf{y}_l|\mathbf{z}_l; \boldsymbol{\theta}) h(\mathbf{z}_l)\, \mathrm{d}\mathbf{z}_l}$$

$$= \sum_{l=1}^{n} \int_{\mathbb{R}^q} S_l(\boldsymbol{\theta}; \mathbf{z}_l) h(\mathbf{z}_l|\mathbf{y}_l; \boldsymbol{\theta})\, \mathrm{d}\mathbf{z}_l = \sum_{l=1}^{n} E_{\mathbf{z}|\mathbf{y}}\big[S_l(\boldsymbol{\theta}; \mathbf{z}_l)\big],$$

where $S_l(\boldsymbol{\theta}; \mathbf{z}_l)$ denotes the complete-data score vector given by $\partial \log f(\mathbf{y}_l, \mathbf{z}_l; \boldsymbol{\theta})/\partial\boldsymbol{\theta} = \partial[\log g(\mathbf{y}_l|\mathbf{z}_l; \boldsymbol{\theta}) + \log h(\mathbf{z}_l)]/\partial\boldsymbol{\theta}$. In words, the observed data score vector is expressed as the expected value of the complete-data vector with respect to $h(\mathbf{z}_l|\mathbf{y}_l; \boldsymbol{\theta})$, that is the posterior distribution of the latent variables given the observations. This implies that (2.2) plays a double role. If the score equations are solved with respect to $\boldsymbol{\theta}$, with $h(\mathbf{z}_l|\mathbf{y}_l; \boldsymbol{\theta})$ fixed at the $\boldsymbol{\theta}$-value of the previous iteration, then this corresponds to the EM algorithm, whereas, if the score equations are solved with respect to $\boldsymbol{\theta}$ considering $h(\mathbf{z}_l|\mathbf{y}_l; \boldsymbol{\theta})$ also as a function of $\boldsymbol{\theta}$, then this corresponds to a direct maximization of the observed data log-likelihood $\ell(\boldsymbol{\theta})$. As we shall discuss further, based on this appealing feature, the estimators derived by applying either of these two algorithms will share the same theoretical properties.

Equation (2.2) involves ratios of multidimensional integrals which cannot be solved analytically, except when all the $g_j(y_{jl}|\mathbf{z}_l; \boldsymbol{\theta})$ are normal. Consequently, an approximation of these integrals is needed, on which the bias and variance of resulting estimators will depend. In this paper, we study the properties of ML estimators based on the adaptive Gauss–Hermite approximation of integrals. This technique consists of adjusting the quadrature locations with specific features of the posterior density of the latent variables given the observations. This provides a better approximation of the function to be integrated. Naylor and Smith [13] took the mean vector and covariance matrix of the normal density approximating the integrand to be the posterior mean and covariance matrix. Unfortunately, these posterior moments are not known exactly, but must themselves be obtained using adaptive quadratures. Integration is therefore iterative. To overcome this limitation, Liu and Pierce [9] proposed an alternative procedure that consists in computing the mode of the integrand and its curvature (inverse of the Hessian matrix) at the mode, so that numerical integration is avoided. In this case, the adaptive quadrature, when applied using one abscissa, is equivalent to the classical Laplace approximation, and its behavior has been analyzed in several papers on generalized linear models (Pinhero and Bates [15], Schilling and Bock [23], Skrondal and Rabe-Hesketh [26], Joe [7]).

The application of the adaptive quadrature requires to rewrite (1.1) as follows

$$f(\mathbf{y}_l; \boldsymbol{\theta}) = \int_{\mathbb{R}^q} \frac{g(\mathbf{y}_l|\mathbf{z}_l; \boldsymbol{\theta}) h(\mathbf{z}_l)}{h_1(\mathbf{z}_l; \hat{\mathbf{z}}_l, \boldsymbol{\Psi}_l)} h_1(\mathbf{z}_l; \hat{\mathbf{z}}_l, \boldsymbol{\Psi}_l)\, \mathrm{d}\mathbf{z}_l, \tag{2.3}$$

where $h_1(\cdot; \hat{\mathbf{z}}_l, \boldsymbol{\Psi}_l)$ is a multivariate normal density with first and second moments

$$\hat{\mathbf{z}}_l = \arg\max_{\mathbf{z}_l \in \mathbb{R}^q} \big[\log g(\mathbf{y}_l|\mathbf{z}_l; \boldsymbol{\theta}) + \log h(\mathbf{z}_l)\big], \tag{2.4}$$

$$\boldsymbol{\Psi}_l = \left(- \frac{\partial^2 [\log g(\mathbf{y}_l|\mathbf{z}_l; \boldsymbol{\theta}) + \log h(\mathbf{z}_l)]}{\partial \mathbf{z}_l^T\, \partial \mathbf{z}_l} \right)\Bigg|_{\mathbf{z}_l = \hat{\mathbf{z}}_l}^{-1}. \tag{2.5}$$

A cartesian product rule based on the classical Gauss–Hermite quadrature is then applied so that the integrals have to be defined with respect to uncorrelated variables $\tilde{\mathbf{z}}_l$. Based on the Cholesky factorization of the covariance matrix $\boldsymbol{\Psi}_l = \mathbf{T}_l \mathbf{T}_l^T$, expression (2.3) can be rewritten as

$$f(\mathbf{y}_l;\boldsymbol{\theta}) = 2^{q/2}|\mathbf{T}_l| \int_{\mathbb{R}^q} g(\mathbf{y}_l|\sqrt{2}\mathbf{T}_l\tilde{\mathbf{z}}_l + \hat{\mathbf{z}}_l;\boldsymbol{\theta}) h(\sqrt{2}\mathbf{T}_l\tilde{\mathbf{z}}_l + \hat{\mathbf{z}}_l) \exp\big[\tilde{\mathbf{z}}_l^T\tilde{\mathbf{z}}_l\big]\exp\big[-\tilde{\mathbf{z}}_l^T\tilde{\mathbf{z}}_l\big]\,\mathrm{d}\tilde{\mathbf{z}}_l,$$

such that the AGH approximation of the density $f(\mathbf{y}_l;\boldsymbol{\theta})$, $l = 1,\ldots,n$, is given by

$$\tilde{f}(\mathbf{y}_l;\boldsymbol{\theta}) = 2^{q/2}|\mathbf{T}_l| \sum_{t_1,\ldots,t_q} g\big(\mathbf{y}_l|\mathbf{z}^*_{l,t_1,\ldots,t_q};\boldsymbol{\theta}\big) h\big(\mathbf{z}^*_{l,t_1,\ldots,t_q}\big) w^*_{t_1}\cdots w^*_{t_q}, \tag{2.6}$$

where $\sum_{t_1,\ldots,t_q} = \sum_{t_1=1}^k \cdots \sum_{t_q=1}^k$, being k the number of quadrature points selected for each latent variable, $\mathbf{z}^*_{l,t_1,\ldots,t_q} = (z^*_{l,t_1},\ldots,z^*_{l,t_q})^T = \sqrt{2}\mathbf{T}_l(z_{t_1},\ldots,z_{t_q})^T + \hat{\mathbf{z}}_l$ and $w^*_{t_k} = w_{t_k}\exp[z^2_{t_k}]$ are the AGH nodes and weights, respectively, with z_{t_k} being the classical GH nodes and w_{t_k}, $k = 1,\ldots,q$, the corresponding weights.

From (2.6), we obtain the approximated log-likelihood function

$$\begin{aligned}\tilde{\ell}(\boldsymbol{\theta}) = \sum_{l=1}^n \log\Bigg[2^{q/2}|\mathbf{T}_l| \sum_{t_1,\ldots,t_q}\prod_{j=1}^p \exp\bigg(&\frac{y_{jl}(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\ldots,t_q}) - b_j(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\ldots,t_q})}{\phi_j} \\ &+ c_j(y_{jl},\phi_j)\bigg) \\ &\times (2\pi)^{-q/2}\exp\bigg(-\frac{1}{2}\mathbf{z}^{*T}_{l,t_1,\ldots,t_q}\mathbf{z}^*_{l,t_1,\ldots,t_q}\bigg) w^*_{t_1}\cdots w^*_{t_q}\Bigg].\end{aligned} \tag{2.7}$$

The estimators of the model parameters are found by equating the corresponding derivatives of (2.7) to zero, that is

$$\begin{aligned}\tilde{S}(\boldsymbol{\theta}) = \frac{\partial\tilde{\ell}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \sum_{l=1}^n \frac{1}{\tilde{f}(\mathbf{y}_l;\boldsymbol{\theta})}\frac{\partial\tilde{f}(\mathbf{y}_l;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \\ &= \sum_{l=1}^n \frac{\sum_{t_1,\ldots,t_q} S_l(\boldsymbol{\theta};\mathbf{z}^*_{l,t_1,\ldots,t_q}) g(\mathbf{y}_l|\mathbf{z}^*_{l,t_1,\ldots,t_q};\boldsymbol{\theta}) h(\mathbf{z}^*_{l,t_1,\ldots,t_q}) w^*_{t_1}\cdots w^*_{t_q}}{\sum_{t_1,\ldots,t_q} g(\mathbf{y}_l|\mathbf{z}^*_{l,t_1,\ldots,t_q};\boldsymbol{\theta}) h(\mathbf{z}^*_{t_1,\ldots,t_q}) w^*_{t_1}\cdots w^*_{t_q}} \\ &= \sum_{l=1}^n \tilde{E}_{\mathbf{z}|\mathbf{y}}\big[S_l(\boldsymbol{\theta};\mathbf{z}_l)\big] = 0,\end{aligned} \tag{2.8}$$

where, specifically,

$$S_l\big(\alpha_{0j};\mathbf{z}^*_{l,t_1,\ldots,t_q}\big) = \frac{1}{\phi_j}\Bigg[y_{jl} - \frac{\partial b_j(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\ldots,t_q})}{\partial\alpha_{0j}}\Bigg],$$

$$S_l(\boldsymbol{\alpha}_j; \mathbf{z}^*_{l,t_1,\dots,t_q}) = \frac{\mathbf{z}^*_{l,t_1,\dots,t_q}}{\phi_j}\left[y_{jl} - \frac{\partial b_j(\alpha_{0j} + \boldsymbol{\alpha}_j^T \mathbf{z}^*_{l,t_1,\dots,t_q})}{\partial \boldsymbol{\alpha}_j}\right]$$

and

$$S_l(\phi_j; \mathbf{z}^*_{l,t_1,\dots,t_q}) = -\frac{1}{\phi_j^2}\left[y_{jl}\left(\alpha_{0j} + \boldsymbol{\alpha}_j^T \mathbf{z}^*_{l,t_1,\dots,t_q}\right) - b_j\left(\alpha_{0j} - \boldsymbol{\alpha}_j^T \mathbf{z}^*_{l,t_1,\dots,t_q}\right)\right] + \frac{\partial c_j(y_{jl}, \phi_j)}{\partial \phi_j}.$$

Equations (2.8) provide a set of estimating equations defining the estimators for the model parameters. The same equations are derived in the E-step of the EM algorithm, in which the AGH quadrature is applied to approximate the E-step expectations (2.2). In the M-step, as in the direct maximization algorithm, improved estimates for the model parameters are obtained by maximizing the approximated expected score functions (2.8). For the scale parameter ϕ_j, closed form expressions can be derived, whereas, for the other parameters, a Newton Raphson iterative scheme is used in order to solve the corresponding nonlinear maximum likelihood equations. In the derivation of the estimating equations, the model has been kept as general as possible without specifying the conditional distributions $g_j(y_j|\mathbf{z}; \boldsymbol{\theta})$. In Appendix C, we give specific expressions for the quantity that are used in the log-likelihood function (2.1) and in the score functions (2.8) for binary manifest variables, whereas we refer to Bianconcini and Cagnone [2] for count and categorical observed variables.

3. Statistical properties of the AGH-based estimators

To investigate the asymptotic properties of the maximum likelihood estimators based on the adaptive Gauss–Hermite quadrature, the error rate associated to the approximation (2.8) has to be determined. Liu and Pierce [9] analyzed the asymptotic behavior of the AGH when it is used to approximate unidimensional integrals. Based on the fact that when applied with only one node it results in the Laplace approximation to the integral (de Bruijn [3], Barndorff-Nielsen and Cox [1]), they proved that the adaptive quadrature based on k points can be alternatively thought as a higher (than one) order Laplace approximation. We now generalize this result to the multidimensional integral (1.1) as well as to the ratio of integrals (2.2), and we analyze the asymptotic accuracy of the corresponding Laplace approximations. The behavior of the latter for multidimensional integrals was studied by Barndorff-Nielsen and Cox [1], Shun and McCullagh [25], Shun [24], and recently by Evangelou, Zhu and Smith [4] for spatial generalized linear mixed models. Similarly, Raudenbush, Yang and Yosef [19] considered improvements of the standard Laplace approximation obtained by incorporating higher order derivatives of the integrand.

For the derivations illustrated here, we follow the notation of Shun and McCullagh [25] based on summation convention. Hence, an index that appears as a subscript and as a superscript implies a summation over all possible values of that index. We will denote the components of a vector sometimes by subscripts and sometimes by superscripts. The (i, j)th component of a matrix $\mathbf{A}$ will be written as a_{ij} and its inverse (when exists) will have components a^{ij}. For any real function $f(\mathbf{z}), \mathbf{z} \in \mathbb{R}^q$, its derivative with respect to the ith component of $\mathbf{z}$ is denoted by a subscript, that is, $f_i(\mathbf{z}) = \frac{\partial f(\mathbf{z})}{\partial z_i}$, $f_{ij}(\mathbf{z}) = \frac{\partial^2 f(\mathbf{z})}{\partial z_i \partial z_j}$, and, more generally, $f_{i_1,\dots,i_{2m}}(\mathbf{z}) = \frac{\partial^{2m} f(\mathbf{z})}{\partial z_{i_1} \cdots \partial z_{i_{2m}}}$. In order to keep the notation as light as possible, we omit the individual subscript l.

3.1. Relationship with the Laplace approximation

The AGH quadrature implemented here is based on a tensor product of q univariate Gaussian quadratures based on the same number of quadrature points. In each dimension, the approximation (2.6) is exact for polynomials of degree $2k+1$ or less. Hence, it provides a good approximation of the integral (1.1) if the ratio $\nu(\mathbf{z}) = \frac{g(\mathbf{y}|\mathbf{z};\boldsymbol{\theta})h(\mathbf{z})}{h_1(\mathbf{z};\hat{\mathbf{z}},\boldsymbol{\Psi})}$ can be approximated well by a q-variate polynomial, where the maximum exponent of all the monomials is at most $2k+1$ (Tauchen and Hussey [27]), in the region where the integrand is substantial. It follows that the effectiveness of the adaptive Gauss–Hermite approximation (2.6) can be evaluated by considering the Taylor series expansion of $\nu(\mathbf{z})$ around the mode $\hat{\mathbf{z}}$, that is,

$$\nu(\mathbf{z}) = \nu(\hat{\mathbf{z}})\left[1 + \sum_{m=3}^{\infty} \frac{1}{m!} c_{i_1,\ldots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\ldots,i_m}\right], \tag{3.1}$$

where $(i_1,\ldots,i_m)$ is a set of m indices, $c_{i_1,\ldots,i_m}(\hat{\mathbf{z}}) = \frac{\nu_{i_1,\ldots,i_m}(\hat{\mathbf{z}})}{\nu(\hat{\mathbf{z}})}$, $\nu_{i_1,\ldots,i_m}(\hat{\mathbf{z}})$ denotes the partial derivatives of order m of ν with respect to $z_{i_1},\ldots,z_{i_m}$ evaluated at the mode $\hat{\mathbf{z}}$, whereas $(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\ldots,i_m}$ refers to specific components of the vector $(\mathbf{z}-\hat{\mathbf{z}})$. The coefficients c_{i_1} and c_{i_1,i_2} are zero due to the choice of $h_1(\cdot;\hat{\mathbf{z}},\boldsymbol{\Psi})$.

Substituting the expansion (3.1) into the integral (1.1), we obtain the exact solution

$$f(\mathbf{y};\boldsymbol{\theta}) = \nu(\hat{\mathbf{z}})\left[1 + \sum_{m=2}^{\infty}\sum_{Q} \frac{1}{(2m)!} c_{i_1,\ldots,i_{2m}}(\hat{\mathbf{z}})\boldsymbol{\nu}^{q_1}(\hat{\mathbf{z}})\cdots\boldsymbol{\nu}^{q_m}(\hat{\mathbf{z}})\right], \tag{3.2}$$

where the second sum is over the partition $Q = q_1|\cdots|q_m$ of $2m$ indices into m blocks, each of size 2, and $\boldsymbol{\nu}^{q_k}(\hat{\mathbf{z}})$, $k=1,\ldots,m$, are components of the covariance matrix $\boldsymbol{\Psi}$. The Gauss–Hermite quadrature, for which k quadrature points are selected for each dimension, would be exact if the partial derivatives beyond the $2(k+1)$ order in (3.2) are zero, that is,

$$f(\mathbf{y};\boldsymbol{\theta}) = \nu(\hat{\mathbf{z}})\left[1 + \sum_{m=2}^{k}\sum_{Q} \frac{1}{(2m)!} c_{i_1,\ldots,i_{2m}}(\hat{\mathbf{z}})\boldsymbol{\nu}^{q_1}(\hat{\mathbf{z}})\cdots\boldsymbol{\nu}^{q_m}(\hat{\mathbf{z}})\right]. \tag{3.3}$$

To determine the asymptotic order of the approximation (3.3), its relationship with the higher order Laplace approximation of multidimensional integrals has to be taken into account. At this regard, the integral (1.1) has to be rewritten as

$$f(\mathbf{y};\boldsymbol{\theta}) = \int_{\mathbb{R}_q} \mathrm{e}^{[-L(\mathbf{z})]}\,\mathrm{d}\mathbf{z}, \tag{3.4}$$

where $L(\mathbf{z}) = -[\log g(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}) + \log h(\mathbf{z})]$, such that $L(\mathbf{z}) = \mathrm{O}(p)$. Assuming that $L(\mathbf{z})$ has a unique minimum $\hat{\mathbf{z}}$, Shun and McCullagh [25] suggested the following expansion around that minimum

$$L(\mathbf{z}) = L(\hat{\mathbf{z}}) + \sum_{m=2}^{\infty} \frac{1}{m!} L_{i_1,\ldots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\ldots,i_m}$$

and applying the exponential function

$$\mathrm{e}^{-L(\mathbf{z})} = (2\pi)^{q/2}|\boldsymbol{\Psi}|^{1/2}\mathrm{e}^{-L(\hat{\mathbf{z}})}h_1(\mathbf{z};\hat{\mathbf{z}},\boldsymbol{\Psi})\exp\left[\sum_{m=3}^{\infty}\frac{(-1)}{m!}L_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}\right],$$

where $h_1(\cdot;\hat{\mathbf{z}},\boldsymbol{\Psi})$ is a multivariate normal density with moments given in (2.4) and (2.5). Based on exlog relations, the higher order term can be expressed as follows

$$\exp\left[\sum_{m=3}^{\infty}\frac{(-1)}{m!}L_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}\right]$$
$$= 1 - \sum_{m=3}^{\infty}\sum_{P}\frac{(-1)^t}{m!}L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_t}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}$$

such that the exact solution of the integral (3.4) is given by

$$(2\pi)^{q/2}|\boldsymbol{\Psi}|^{1/2}\mathrm{e}^{-L(\hat{\mathbf{z}})}\left[1 - \sum_{m=2}^{\infty}\sum_{P,Q}\frac{(-1)^t}{(2m)!}L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_t}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}})\right], \tag{3.5}$$

where the second sum is over all partitions P, Q, such that $P = p_1|\cdots|p_t$ is a partition of $2m$ indices into t blocks, each of size 3 or more, and $Q = q_1|\cdots|q_m$ is a partition of $2m$ indices into m blocks, each of size 2. Each component $L^{q_k}(\hat{\mathbf{z}})$, $k = 1,\dots,m$, refers to specific elements of $\boldsymbol{\Psi}$. As shown in the Appendix A, the exact solution (3.5) is equivalent to the one derived in (3.2). It follows that the asymptotic order of the AGH approximation can be derived by truncating at $m = k$ the expansion (3.5), and by analyzing the asymptotic order associated to the bipartition (P, Q) related to $m = k + 1$. For fixed q, the usual asymptotic order of the term corresponding to the bipartition (P, Q) in (3.5) is $\mathrm{O}(p^{t-m})$. It follows that the asymptotic error of the AGH based on k quadrature points is the same associated to the bipartition (P, Q) of $2(k + 1)$ indices, that is, $\mathrm{O}(p^{-[k/3+1]})$ (see in Appendix A for more details).

It has to be noticed that when AGH quadratures are applied in the estimation of GLLVM, we need to approximate ratios of integrals as shown in (2.2). The fully exponential solution (3.5) cannot be applied to the integral at the numerator, since the score functions $S(\boldsymbol{\theta};\mathbf{z})$ are not necessarily positive. The integral has to be written in the standard form (Tierney, Kass and Kadane [28], Evangelou, Zhu and Smith [4])

$$\int_{\mathbb{R}^q}\mathrm{e}^{-L(\mathbf{z})}S(\boldsymbol{\theta};\mathbf{z})\,\mathrm{d}\mathbf{z}.$$

Beyond the Taylor series expansion of $L(\mathbf{z})$ around its minimum $\hat{\mathbf{z}}$, we have to consider a similar expansion of S around the same point, that is,

$$S(\boldsymbol{\theta};\mathbf{z}) = \sum_{m=0}^{\infty}S_{j_1,\dots,j_m}(\boldsymbol{\theta};\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{j_1,\dots,j_m}.$$

Following Evangelou, Zhu and Smith [4], it can be shown that

$$\int_{\mathbb{R}^q} S(\boldsymbol{\theta};\mathbf{z})g(\mathbf{y}|\mathbf{z};\boldsymbol{\theta})h(\mathbf{z})\,\mathrm{d}\mathbf{z}$$
$$= (2\pi)^{q/2}|\boldsymbol{\Psi}|^{1/2}\mathrm{e}^{-[L(\hat{\mathbf{z}})]}$$
$$\times\left[\sum_{m=0}^{\infty}\sum_{s=0}^{2m}\sum_{P,Q}\frac{(-1)^t}{(2m)!}S_{j_1,\dots,j_s}(\boldsymbol{\theta};\hat{\mathbf{z}})L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_t}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}})\right],$$

where P is a partition of $2m-s$ indices into t blocks, each of size 3 or more, and Q is a partition of the same indices together with $\{j_1,\dots,j_s\}$ into m blocks of size 2. Note that P and Q do not need to be connected. It follows that the exact Laplace solution of the expected score function (2.2) results

$$\frac{\sum_{m=0}^{\infty}\sum_{s=0}^{2m}\sum_{P,Q}((-1)^t/(2m!))S_{j_1,\dots,j_s}(\boldsymbol{\theta};\hat{\mathbf{z}})L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_t}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}})}{\sum_{m=0}^{\infty}\sum_{P,Q}((-1)^t/(2m!))L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_t}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}})}. \tag{3.6}$$

It will be perfectly account for the AGH approximation in (2.8) if the partial derivatives, at both the numerator and denominator, of order greater than $2k$ ($\max m = k$) are zero. The corresponding Laplace approximation can be rewritten by regrouping in decreasing asymptotic order the elements that appear in both the expansions, and by truncating the resulting series at an appropriate point. In symbols,

$$\frac{S(\boldsymbol{\theta};\hat{\mathbf{z}})+c_1^*p^{-1}+\cdots+c_r^*p^{-r}+\cdots+c_{[k/3]}^*p^{-[k/3]}+\mathrm{O}(p^{-[k/3+1]})}{1+c_1p^{-1}+\cdots+c_rp^{-r}+\cdots+c_{[k/3]}p^{-[k/3]}+\mathrm{O}(p^{-[k/3+1]})}, \tag{3.7}$$

where the coefficients $c_r, r=1,\dots,[\frac{k}{3}]$, are given by

$$c_r=\sum_{m=r+1}^{3r}\frac{(-1)^{m-r}}{(2m)!}L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_{m-r}}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}})$$

with $p_1|\cdots|p_t$ be a partition of $2m$ indices into $m-r$ blocks, each of size 3 or more, and $q_1|\cdots|q_m$ is a partition of $2m$ indices into m blocks, each of size 2. On the other hand, the coefficients $c_r^*, r=1,\dots,[\frac{k}{3}]$, results

$$c_r^*=\sum_{m=r}^{3r}\sum_{s=0}^{3r-m}\frac{(-1)^{m-r}}{(2m)!}S_{j_1,\dots,j_s}(\boldsymbol{\theta};\hat{\mathbf{z}})L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_{m-r}}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}}),$$

where $p_1|\cdots|p_{m-r}$ is a partition of $2m-s$ indices into $m-r$ blocks, each of size 3 or more, and $q_1|\cdots|q_m$ is a partition of the same indices together with $\{j_1,\dots,j_s\}$ into m blocks of size 2. Since $S_{j_1,\dots,j_s}(\boldsymbol{\theta};\hat{\mathbf{z}})=\frac{\partial L_{j_1,\dots,j_s}(\hat{\mathbf{z}})}{\partial\boldsymbol{\theta}}$, all the first derivatives of the score function will be zero due to the choice of $\hat{\mathbf{z}}$.

Based on long polynomial division, the approximated expected score functions equivalent to (2.8) are given by

$$\sum_{l=1}^{n} \tilde{E}_{\mathbf{z}|\mathbf{y}}\big[S_l(\boldsymbol{\theta}; \mathbf{z}_l)\big] = \sum_{l=1}^{n} \big[S_l(\boldsymbol{\theta}; \hat{\mathbf{z}}_l) + c_1^{**} p^{-1} + \cdots + c_r^{**} p^{-r} + \cdots + \mathrm{O}\big(p^{-[k/3+1]}\big)\big], \tag{3.8}$$

where the coefficients $c_r^{**}, r = 1, \ldots, [\frac{k}{3}]$, can be determined as follows

$$c_r^{**} = \big[c_r^* - S(\boldsymbol{\theta}; \hat{\mathbf{z}})c_r\big] - c_{r-1}c_1^* - c_{r-2}c_2^* - \cdots - c_1 c_{r-1}^*$$

being

$$c_r^* - S(\boldsymbol{\theta}; \hat{\mathbf{z}})c_r = \sum_{m=r}^{3r} \sum_{s=2}^{3r-m} \frac{(-1)^{m-r}}{(2m)!} S_{j_1,\ldots,j_s}(\boldsymbol{\theta}; \hat{\mathbf{z}}) L_{p_1}(\hat{\mathbf{z}}) \cdots L_{p_{m-r}}(\hat{\mathbf{z}}) L^{q_1}(\hat{\mathbf{z}}) \cdots L^{q_m}(\hat{\mathbf{z}})$$

and, in particular, $c_1^* - S(\boldsymbol{\theta}; \hat{\mathbf{z}})c_1 = \frac{1}{2} S_{j_1,j_2} L^{j_1,j_2}(\hat{\mathbf{z}})$.

3.2. Asymptotic behavior of the AGH-based estimators

To investigate the properties of the AGH approximated maximum likelihood estimators $\hat{\boldsymbol{\theta}}$, we analyze the asymptotic behavior of the corresponding Laplace-based estimators defined by (3.8). Our arguments are similar to those of Huber, Ronchetti and Victoria-Feser [5], who discussed classical Laplace estimators in GLLVM, and Rizopoulos, Verbeke and Lesaffre [20] who analyzed the consistency of fully exponential Laplace estimators in joint models for survival and longitudinal data.

Proposition 3.2.1 (Consistency). *Let $\boldsymbol{\theta}_0 \in \Theta$ denote the true parameter value, then, under suitable regularity conditions,*

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathrm{O}_p\big[\max\big(n^{-1/2}, p^{-[k/3+1]}\big)\big]. \tag{3.9}$$

Thus, $\hat{\boldsymbol{\theta}}$ is consistent as long as both n and p grow to ∞. A formal proof of Proposition 3.2.1 is given in Appendix B. The $n^{-1/2}$ term comes from the standard asymptotic theory, whereas the $p^{-[k/3+1]}$ term derives from the AGH approximation. The requirement that p grows to infinity is consistent with the fact that we are trying to approximate the marginal density of each individual, that is, $f(\mathbf{y}_l; \boldsymbol{\theta})$. However, in practical applications where p and k are both fixed, the approximation error in the adaptive technique is $\mathrm{O}(p^{-[k/3+1]})$ as $n \to \infty$, and the asymptotic properties of the AGH-based estimators should be evaluated with respect to a perturbation of the true parameter $\boldsymbol{\theta}_0$.

For $k \geq 3$, the AGH-based estimator is more accurate than the classical $\mathrm{O}(p^{-1})$ Laplace-based estimators. Indeed, it shares the same accuracy of higher (than one) order Laplace estimators, but, with respect to this latter, the adaptive Gauss–Hermite is easier to be implemented, since it avoids derivative computations.

Based on the derivation of (3.9) as presented in the Appendix B, we can deduce that, if $p = \mathrm{O}(n^{\rho})$ for $\rho > \frac{1}{[k/3+1]}$, then the AGH-based estimators will be asymptotically equivalent to the true maximum likelihood estimators that solve $S(\boldsymbol{\theta}) = 0$. However, in general, they are not maximum likelihood estimators because of the approximation, but, as discussed by Huber, Ronchetti and Victoria-Feser [5] for classical Laplace estimators in GLLVM, they belong to the class of M-estimators. The latter are implicitly defined through a general Ψ-function as the solution in $\boldsymbol{\theta}$ of

$$\sum_{l=1}^{n} \Psi(\mathbf{y}_l; \boldsymbol{\theta}) = 0.$$

The Ψ-function for the AGH-based estimators are given by (2.8).

Proposition 3.2.2 (Asymptotic normality). *If $\boldsymbol{\theta}_0$ is an interior point of the parameter space Θ and $B(\boldsymbol{\theta}_0) = -E[\frac{\partial \Psi(\mathbf{y}_l;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}] = -E[\frac{\partial^2 \tilde{\ell}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T}]$ is nonsingular, the AGH-based estimators are asymptotically normal, that is,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to^{D} MVN\big(\mathbf{0}, B(\boldsymbol{\theta}_0)^{-1} A(\boldsymbol{\theta}_0)\big[B(\boldsymbol{\theta}_0)^{-1}\big]^T\big) \tag{3.10}$$

with $A(\boldsymbol{\theta}_0) = E[\Psi(\mathbf{y}_l; \boldsymbol{\theta}_0)\Psi^T(\mathbf{y}_l; \boldsymbol{\theta}_0)] = E[\frac{\partial \tilde{\ell}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\ell}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}^T]$.

The regularity conditions that ensure consistency and asymptotic normality of the AGH-based M-estimators have to be checked for the particular conditional distribution of each y_j (Huber, Ronchetti and Victoria-Feser [5]). For classical Laplace-based estimators, Huber, Scaillet and Victoria-Feser [6] analyzed these conditions for ordered multinomial distributed manifest variables. A formal derivation for the M-estimators discussed here in the case of binary observed variables is provided in Appendix C.

4. Monte Carlo simulations

In this section, we investigate empirically the finite sample performance of the adaptive Gauss–Hermite and related Laplace-based estimators. We focus on latent variable models for binary data, since in this case the differences between numerical techniques should be better highlighted (Joe [7]). We consider two simulation scenarios characterized by an increasing number of observed and latent variables. In particular, we generate data from a population that consists of six items satisfying a three factor model, and from a population based on ten observed variables that satisfy a five factor model. In both cases, the population parameters have been chosen in such a way that the item-specific intercepts and the factor loadings are drawn randomly from a log-normal distribution, with some loadings fixed to 0 to get unique solutions. For each scenario, 100 random samples have been considered with 200 subjects.

A crucial choice in the application of the AGH quadrature is the number of points needed to adequately approximate the likelihood function. In the simulation study, we follow Schilling and Bock [23] who suggested to select, in presence of binary data, five and three quadrature points for the three and five factor model, respectively. In both cases, the performance of AGH is compared with that of the Laplace approximation of order $O(p^{-2})$. The estimation is performed through the direct maximization algorithm described in Section 2, whose mathematical details for the case of binary observed items are provided in Appendix C. The algorithm is written in the statistical language R (R Development Core Team [16]) and the program is available from the authors on request.

In the case of each simulation, the true values used to generate the samples, the mean values of the estimated parameters across simulations, together with their corresponding standard deviations obtained from the simulated results, the mean estimated standard errors obtained from (3.10) and the Root Mean Square Error (RMSE) are reported. Furthermore, in order to better highlight the computational burden of each technique under the different conditions of study, we report the average (over all the generated samples) computational time in minutes (Avg min) and the average number of iterations (Avg iter) required by the algorithm to get the convergence in a sample (obtain on Intel Core i7 quad-core, 3.1 GHz CPU with 16 Gb RAM).

Table 1 studies the performance of the AGH based on five quadrature points and of the second order Laplace approximation on the data generated by the three factor model. The results

Table 1. True values, mean, simulated standard deviations (S.D.), root mean square error (RMSE) and estimated standard errors (S.E.) of the parameter estimates for AGH based on 5 quadrature points and for second order Laplace (*Lap*2) approximation in data generated by a three factor model with six ($p = 6$) items observed on $n = 200$ subjects

	AGH				*Lap2*			
True	Mean	S.D.	RMSE	S.E.	Mean	S.D.	RMSE	S.E.
$\alpha_{11} = 1.01$	0.72	0.37	0.47	0.19	1.32	0.70	0.77	0.19
$\alpha_{21} = 0.91$	1.17	0.36	0.45	0.46	1.11	0.38	0.42	0.49
$\alpha_{31} = 0.50$	0.39	0.31	0.34	0.16	0.35	0.34	0.37	0.36
$\alpha_{41} = 0.74$	0.99	0.38	0.45	0.27	1.14	0.18	0.44	0.25
$\alpha_{51} = 1.16$	1.39	0.37	0.44	0.57	1.68	0.37	0.64	0.66
$\alpha_{61} = 1.22$	1.54	0.44	0.55	0.26	1.23	0.52	0.52	0.42
$\alpha_{12} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{22} = 0.83$	0.45	0.30	0.49	0.32	0.21	0.69	0.93	0.54
$\alpha_{32} = 0.44$	1.02	0.38	0.69	0.42	1.06	0.36	0.71	0.61
$\alpha_{42} = 0.88$	1.15	0.42	0.50	0.57	1.13	0.48	0.54	0.53
$\alpha_{52} = 1.73$	2.54	0.53	0.96	0.91	2.53	0.37	0.88	0.85
$\alpha_{62} = 1.46$	1.43	0.52	0.52	0.46	1.36	0.73	0.73	0.53
$\alpha_{13} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{23} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{33} = 1.45$	1.08	0.44	0.58	0.49	1.21	0.68	0.72	0.69
$\alpha_{43} = 1.05$	1.52	0.42	0.64	0.62	1.49	0.36	0.57	0.47
$\alpha_{53} = 0.62$	0.93	0.36	0.48	0.58	0.80	0.37	0.41	0.66
$\alpha_{63} = 0.91$	0.98	0.42	0.43	0.42	0.53	0.34	0.51	0.41
Avg iter	9.21				324.94			
Avg min	$3'52''$				$24'26''$			

show that the two techniques provide similar RMSE values for almost all the model parameters. Indeed, even if the Laplace seems to introduce a slightly larger bias in some estimates than the AGH, the simulated standard deviations, that are a measure of the sampling variability of the estimated parameters, are quite close. The estimated standard errors are generally larger than the simulated standard deviations for both the techniques, and closer to the corresponding RMSE. As expected, the main difference between the two techniques is computational. The algorithm based on the adaptive quadrature achieves convergence for a sample, on average, in ten iterations, that is in less than four minutes, whereas the second order Laplace requires, on average, more than 300 iterations to get the convergence in a sample, that means almost thirty minutes. This is more evident in Table 2 that shows the results for the five factor model. In this specific case, the adaptive Gauss–Hermite has been applied with three quadrature points. For this latter,

Table 2. True values, mean, simulated standard deviations (S.D.), root mean square error (RMSE) and estimated standard errors (S.E.) of the parameter estimates for AGH based on three quadrature points, and for second order Laplace (*Lap*2) approximation in the data generated by a five factor model with ten ($p = 10$) items observed on $n = 200$ individuals

	AGH				*Lap2*			
True	Mean	S.D.	RMSE	S.E.	Mean	S.D.	RMSE	S.E.
$\alpha_{11} = 1.01$	0.70	0.34	0.46	0.64	1.28	0.46	0.53	0.54
$\alpha_{21} = 0.91$	1.27	0.27	0.45	0.56	1.33	0.51	0.66	0.56
$\alpha_{31} = 0.50$	0.87	0.41	0.55	0.52	0.57	0.22	0.23	0.42
$\alpha_{41} = 0.74$	1.14	0.39	0.56	0.61	0.85	0.33	0.35	0.58
$\alpha_{51} = 1.16$	1.83	0.26	0.72	0.71	1.98	0.48	0.95	0.81
$\alpha_{61} = 1.22$	1.22	0.39	0.39	0.42	0.66	0.23	0.60	0.62
$\alpha_{71} = 0.55$	0.48	0.31	0.32	0.27	0.59	0.22	0.22	0.27
$\alpha_{81} = 0.83$	1.10	0.35	0.44	0.45	0.90	0.26	0.27	0.35
$\alpha_{91} = 0.44$	1.01	0.28	0.63	0.64	1.05	0.21	0.65	0.64
$\alpha_{101} = 0.88$	1.05	0.30	0.35	0.36	1.17	0.18	0.34	0.43
$\alpha_{12} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{22} = 1.46$	1.39	0.26	0.26	0.26	1.21	0.44	0.50	0.28
$\alpha_{32} = 0.89$	0.59	0.50	0.58	0.57	0.80	0.38	0.39	0.53
$\alpha_{42} = 1.64$	1.27	0.23	0.44	0.45	1.37	0.38	0.47	0.52
$\alpha_{52} = 1.45$	0.60	0.35	0.92	0.91	0.59	0.46	0.98	0.91
$\alpha_{62} = 1.05$	0.92	0.37	0.39	0.38	1.06	0.19	0.19	0.28
$\alpha_{72} = 0.62$	0.68	0.42	0.42	0.43	0.80	0.34	0.39	0.43
$\alpha_{82} = 0.91$	0.40	0.32	0.60	0.58	0.19	0.41	0.83	0.60
$\alpha_{92} = 1.59$	1.22	0.31	0.48	0.48	2.02	0.59	0.73	0.52
$\alpha_{102} = 1.27$	0.95	0.32	0.46	0.46	1.22	0.27	0.27	0.39
$\alpha_{13} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{23} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{33} = 0.71$	1.10	0.45	0.59	0.62	1.13	0.51	0.66	0.65
$\alpha_{43} = 0.35$	1.02	0.29	0.73	0.74	1.01	0.15	0.68	0.65
$\alpha_{53} = 0.53$	1.46	0.28	0.97	0.98	1.46	0.16	0.95	0.98

Table 2. (Continued)

	AGH				*Lap2*			
True	Mean	S.D.	RMSE	S.E.	Mean	S.D.	RMSE	S.E.
$\alpha_{63} = 0.83$	0.97	0.40	0.42	0.45	0.64	0.25	0.31	0.41
$\alpha_{73} = 0.71$	1.12	0.36	0.55	0.56	1.52	0.37	0.69	0.59
$\alpha_{83} = 0.65$	1.36	0.30	0.77	0.77	1.27	0.47	0.78	0.77
$\alpha_{93} = 0.95$	1.19	0.30	0.39	0.41	0.84	0.17	0.21	0.39
$\alpha_{103} = 0.88$	1.23	0.38	0.52	0.54	0.68	0.18	0.37	0.39
$\alpha_{14} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{24} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{34} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{44} = 1.10$	1.42	0.37	0.49	0.51	1.95	0.40	0.65	0.54
$\alpha_{54} = 0.50$	0.84	0.29	0.45	0.46	0.95	0.57	0.53	0.49
$\alpha_{64} = 0.49$	0.93	0.42	0.61	0.62	0.05	0.61	0.56	0.62
$\alpha_{74} = 1.20$	0.67	0.51	0.73	0.74	0.50	0.24	0.74	0.74
$\alpha_{84} = 0.41$	0.43	0.36	0.36	0.36	0.11	0.31	0.43	0.38
$\alpha_{94} = 0.85$	0.82	0.37	0.37	0.41	0.43	0.20	0.47	0.40
$\alpha_{104} = 0.72$	1.03	0.37	0.48	0.48	1.11	0.21	0.44	0.42
$\alpha_{15} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{25} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{35} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{45} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{55} = 0.62$	0.80	0.28	0.33	0.34	0.53	0.35	0.36	0.34
$\alpha_{65} = 0.99$	1.32	0.32	0.47	0.48	1.23	0.62	0.57	0.52
$\alpha_{75} = 1.12$	1.04	0.36	0.37	0.38	0.97	0.36	0.35	0.40
$\alpha_{85} = 0.86$	1.05	0.40	0.45	0.46	1.42	0.52	0.37	0.43
$\alpha_{95} = 0.71$	0.67	0.33	0.33	0.36	0.94	0.22	0.32	0.35
$\alpha_{105} = 1.39$	1.32	0.35	0.36	0.35	1.73	0.54	0.25	0.35
Avg iter	9.15				344.27			
Avg min	$4'47''$				$239'16''$			

the algorithm requires, on average, less than ten iterations to get convergence in a sample, that is less than five minutes, whereas the algorithm based on the second order Laplace approximation is much slower than in the case of the three factor model. As before, it reaches convergence, on average, in almost 350 iterations, but now it requires almost four hours to obtain the solution for one sample. However, the estimates derived by applying the two techniques are quite comparable in terms of bias, standard deviations and RMSE, with similar conclusions to those drawn for the first scenario.

To better investigate the properties of the adaptive ML estimators, a further simulation study has been conducted in order to understand how much contribution is due to the approximation error and how much is due to the term $O(n^{1/2})$ in the rate of consistency (3.9). For the three factor model, the performance of the adaptive quadrature based on five quadrature points (*AGH*5) has

Table 3. True values, mean, simulated standard deviations (S.D.), root mean square error (RMSE) and estimated standard errors (S.E.) of the parameter estimates for AGH based on 5 (*AGH*5), 9 (*AGH*9) and 15 (*AGH*15) quadrature points in data generated by a three factor model with six observed items ($p = 6$)

	AGH5							
	$n = 200$				$n = 1000$			
True	Mean	S.D.	RMSE	S.E.	Mean	S.D.	RMSE	S.E.
$\alpha_{11} = 1.01$	0.72	0.37	0.47	0.19	0.89	0.12	0.17	0.17
$\alpha_{21} = 0.91$	1.17	0.36	0.45	0.46	0.98	0.06	0.10	0.19
$\alpha_{31} = 0.50$	0.39	0.31	0.34	0.16	0.53	0.00	0.03	0.08
$\alpha_{41} = 0.74$	0.99	0.38	0.45	0.27	0.90	0.00	0.16	0.15
$\alpha_{51} = 1.16$	1.39	0.37	0.44	0.57	1.35	0.00	0.19	0.19
$\alpha_{61} = 1.22$	1.54	0.44	0.55	0.26	1.31	0.22	0.24	0.26
$\alpha_{12} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{22} = 0.83$	0.45	0.30	0.49	0.32	0.73	0.00	0.11	0.12
$\alpha_{32} = 0.44$	1.02	0.38	0.69	0.42	0.87	0.00	0.42	0.31
$\alpha_{42} = 0.88$	1.15	0.42	0.50	0.57	1.07	0.00	0.19	0.17
$\alpha_{52} = 1.73$	2.54	0.53	0.96	0.91	2.10	0.62	0.75	0.71
$\alpha_{62} = 1.46$	1.43	0.52	0.52	0.46	1.45	0.28	0.28	0.26
$\alpha_{13} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{23} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{33} = 1.45$	1.08	0.44	0.58	0.49	1.19	0.31	0.41	0.39
$\alpha_{43} = 1.05$	1.52	0.42	0.64	0.62	1.27	0.38	0.45	0.44
$\alpha_{53} = 0.62$	0.93	0.36	0.48	0.58	0.80	0.35	0.39	0.38
$\alpha_{63} = 0.91$	0.98	0.42	0.43	0.42	0.90	0.29	0.35	0.32
Avg iter	9.21				11.66			
Avg min	$3'52''$				$41'28''$			

been analyzed in presence of small ($n = 200$) and large ($n = 1000$) samples. Furthermore, for the smallest sample size, the behavior of AGH has been studied by also considering nine (*AGH*9) and fifteen (*AGH*15) quadrature points. All the results are illustrated in Table 3. We recall that *AGH*9 shares the same asymptotic properties of the Laplace estimator of order $O(p^{-4})$, whereas when fifteen quadrature points are used, the error rate is of order $O(p^{-6})$. However, the performance of these Laplace estimators is not analyzed since they require a lot of time just to run a simple simulation example as the one considered here.

In Table 3, it can be noticed that *AGH*5 performs better in the largest sample than in the smallest one in terms of both bias and RMSE. The estimated standard errors for all the parameters become smaller and also closer to the root mean square error as the sample size increases.

On the other hand, increasing the number of quadrature points, the AGH performs better in terms of bias and RMSE with slight differences between *AGH*9 and *AGH*15, mainly due to a less variability in the estimates for the latter than for the former. However, the adaptive Gauss–Hermite quadrature becomes more computational intensive as the number of quadrature points increases. As shown in Table 3, the algorithm needs, on average, almost ten iterations to get

Table 3. (Continued)

	AGH9				*AGH15*			
	$n = 200$				$n = 200$			
True	Mean	S.D.	RMSE	S.E.	Mean	S.D.	RMSE	S.E.
$\alpha_{11} = 1.01$	0.82	0.32	0.37	0.17	0.83	0.28	0.31	0.29
$\alpha_{21} = 0.91$	1.01	0.33	0.34	0.44	1.00	0.24	0.26	0.35
$\alpha_{31} = 0.50$	0.39	0.32	0.34	0.38	0.60	0.24	0.25	0.28
$\alpha_{41} = 0.74$	0.94	0.40	0.45	0.28	0.93	0.32	0.32	0.27
$\alpha_{51} = 1.16$	1.29	0.36	0.38	0.54	1.38	0.32	0.39	0.39
$\alpha_{61} = 1.22$	1.56	0.35	0.49	0.30	1.48	0.41	0.44	0.32
$\alpha_{12} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{22} = 0.83$	0.55	0.31	0.42	0.43	0.63	0.24	0.27	0.32
$\alpha_{32} = 0.44$	0.93	0.44	0.65	0.51	0.89	0.18	0.42	0.43
$\alpha_{42} = 0.88$	1.04	0.42	0.45	0.41	1.06	0.25	0.26	0.31
$\alpha_{52} = 1.73$	2.40	0.45	0.80	0.77	2.22	0.50	0.70	0.70
$\alpha_{62} = 1.46$	1.43	0.50	0.50	0.33	1.43	0.48	0.48	0.43
$\alpha_{13} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{23} = 0.00$	–	–	–	–	–	–	–	–
$\alpha_{33} = 1.45$	1.10	0.36	0.50	0.51	1.12	0.40	0.50	0.47
$\alpha_{43} = 1.05$	1.46	0.37	0.56	0.52	1.42	0.33	0.49	0.46
$\alpha_{53} = 0.62$	0.94	0.40	0.51	0.49	0.89	0.41	0.45	0.45
$\alpha_{63} = 0.91$	0.99	0.43	0.43	0.42	0.95	0.45	0.46	0.46
Avg iter	9.40				12.74			
Avg min	$19'\,37''$				$107'\,11''$			

convergence in presence of both five and nine quadrature points. However, whereas in the former case, the solution for a sample is obtained, on average, in four minutes, more than fifteen minutes are required in the latter case. This is more evident for *AGH*15, for which the algorithm gets convergence, on average, in almost thirteen iterations, but requiring more than one hour and a half to obtain the solution for one sample. It is also evident that to get the same accuracy in the estimates observed for *AGH*5 in the largest sample, in presence of binary data, more than fifteen quadrature points per dimension should be considered in small samples.

5. Conclusions

In this paper, we have investigated the theoretical properties of adaptive Gauss–Hermite based estimators in the GLLVM framework. Recently, the adaptive quadrature has played a prominent role in the latent variable model literature for approximating integrals defined over the latent space. It allows to overcome the main limitations of the commonly used techniques, such as the Gauss–Hermite quadrature and the standard Laplace approximation. Indeed, AGH is applicable to problems involving high-dimensional integrals where the former becomes impractical or computationally intensive, and it provides more accurate estimates than the latter, particularly when used for binary or ordinal data with small sample sizes (Joe [7]).

We have proved that, for multidimensional integrals, the AGH solution is asymptotically equivalent to the Laplace approximation that involves specific higher (than two) order derivatives of the integrand. Higher order Laplace approximations have been suggested in several papers on generalized linear models (Raudenbush, Yang and Yosef [19], Evangelou, Zhu and Smith [4], Bianconcini and Cagnone [2]) as an alternative to classical methods for improving the accuracy of the estimates. This extension has been motivated by the well-known asymptotic properties that characterize the Laplace method, and by the fact that the approach does not suffer from the "curse of dimensionality". However, the inclusion of higher order terms is computationally demanding as the order of the approximation increases. On the other hand, the AGH quadrature is easier to be implemented, but of course its computational complexity increases as the number of latent variables increases. Hence, AGH and higher order Laplace approximations can be seen as complementary approaches that share the same asymptotic properties.

We have shown that the AGH-based estimators are consistent as the sample size and number of observed variables grow to infinity. The convergence rate of these estimators depends also on the number of quadrature points used for each dimension. In general, these estimators are less efficient than maximum likelihood estimators because of the approximation, but belong to the class of M-estimators, for which the asymptotic properties are well-known such that correct inference can be performed.

Appendix A: Asymptotic behavior of the multivariate AGH approximation

The higher order Laplace approximation of (1.1) is derived by considering

$$f(\mathbf{y};\boldsymbol{\theta}) = \int_{\mathbb{R}^q} \mathrm{e}^{-L(\mathbf{z})}\,\mathrm{d}\mathbf{z},$$

where $L(\mathbf{z}) = -[\log g(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}) + \log h(\mathbf{z})]$, being $L(\mathbf{z}) = \mathrm{O}(p)$. It is based on the Taylor series expansion of L around its minimum $\hat{\mathbf{z}}$, that is,

$$L(\mathbf{z}) = L(\hat{\mathbf{z}}) + \frac{1}{2}L_{i_1,i_2}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,i_2} + \sum_{m=3}^{\infty}\frac{1}{m!}L_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}. \tag{A.1}$$

Substituting (A.1) into the integral, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^q} \exp\bigl[-L(\mathbf{z})\bigr]\,\mathrm{d}\mathbf{z} \\
&\quad= (2\pi)^{q/2}|\boldsymbol{\Psi}|^{1/2}\mathrm{e}^{-L(\hat{\mathbf{z}})}\int_{\mathbb{R}^q} h_1(\mathbf{z};\hat{\mathbf{z}},\boldsymbol{\Psi})\exp\Biggl[\sum_{m=3}^{\infty}\frac{(-1)}{m!}L_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}\Biggr]\mathrm{d}\mathbf{z} \\
&\quad= (2\pi)^{q/2}|\boldsymbol{\Psi}|^{1/2}\mathrm{e}^{-L(\hat{\mathbf{z}})}\Biggl[1 - \sum_{m=2}^{\infty}\sum_{P,Q}\frac{(-1)^t}{(2m)!}L_{p_1}(\hat{\mathbf{z}})\cdots L_{p_t}(\hat{\mathbf{z}})L^{q_1}(\hat{\mathbf{z}})\cdots L^{q_m}(\hat{\mathbf{z}})\Biggr],
\end{aligned}$$

where the second sum is over all partitions P, Q, such that $P = p_1|\cdots|p_t$ is a partition of $2m$ indices into t blocks, each of size 3 or more, and $Q = q_1|\cdots|q_m$ is a partition of $2m$ indices into m blocks, each of size 2. Each component $L^{q_k}(\hat{\mathbf{z}}), k = 1, \dots, m$, refers to specific elements of the covariance matrix $\boldsymbol{\Psi}$.

We want here to show that the exact higher order Laplace solution for the integral (1.1) is equivalent to the one based on the AGH quadrature given in (3.2). To do so, we need to show that

$$\begin{aligned} 1 + \sum_{m=3}^{\infty} \sum_{P} (-1)^t L_{p_1}(\hat{\mathbf{z}}) \cdots L_{p_t}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m} \\ = \sum_{m=3}^{\infty} c_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}. \end{aligned} \tag{A.2}$$

At this regard, we can notice that, based on the exlog relations, the LHS term of (A.2) is equal to $\exp[\sum_{m=3}^{\infty} L_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}]$. This higher order term can be rewritten as

$$\pi(\mathbf{z}) = \exp\left[-L(\mathbf{z}) + L(\hat{\mathbf{z}}) + \tfrac{1}{2} L_{i_1,i_2}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,i_2}\right]$$

that it is equal to

$$\frac{(2\pi)^{-q/2}|\boldsymbol{\Psi}|^{-1/2} g(\mathbf{y}|\mathbf{z};\boldsymbol{\theta})h(\mathbf{z})}{g(\mathbf{y}|\hat{\mathbf{z}};\boldsymbol{\theta})h(\hat{\mathbf{z}})h_1(\mathbf{z};\hat{\mathbf{z}},\boldsymbol{\Psi})} = \frac{\nu(\mathbf{z})}{\nu(\hat{\mathbf{z}})} = c(\mathbf{z}).$$

Hence, the Taylor series expansion of $\pi(\mathbf{z})$ around the minimum $\hat{\mathbf{z}}$ can be written as

$$\begin{aligned} \pi(\mathbf{z}) &= 1 + \sum_{m=3}^{\infty} \frac{1}{m!} \pi_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m} \\ &= 1 + \sum_{m=3}^{\infty} \frac{1}{m!} c_{i_1,\dots,i_m}(\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{i_1,\dots,i_m}. \end{aligned}$$

It follows that the AGH solution (3.2) and the Laplace one (3.5) are equivalent. Based on this relationship, it is possible to derive the asymptotic error associated with the AGH approximation (3.3) evaluating the equivalent Laplace approximation obtained by truncating (3.5) at $m = k$. Shun and McCullagh [25] proved that, for fixed q, the usual asymptotic order of the term corresponding to the bipartition (P, Q) is $\mathrm{O}(p^{t-m})$. The error rate of the AGH based on k quadrature points is the same associated to the bipartition (P, Q) of $2(k+1)$ indices in the expansion (3.5). In this case, the maximum number of blocks, each of size at least 3, for $2(k+1)$ indices is $[\frac{2(k+1)}{3}]$, where $[r]$ indicates the largest integer not exceeding r. Hence, being $m = k+1$, the AGH based on k quadrature points has associated asymptotic order equal to $\mathrm{O}(p^{-[k/3+1]})$.

Appendix B: Consistency of the AGH-based estimators

This section concerns with the consistency of the AGH-based estimators. All the following proofs proceed along the lines of Vonesh [29], who derived the rate of convergence of the estimator based on the classical Laplace approximation for nonlinear mixed effect models, and of Rizopoulos, Verbeke and Lesaffre [20], who derived that rate for fully exponential Laplace based estimators in joint models for longitudinal and survival data. In particular, we work under the following assumptions:

1. $\hat{\mathbf{z}} = \arg\max_{z\in\mathbb{R}^q}[\log g(\mathbf{y}|\mathbf{z};\boldsymbol{\theta}) + \log h(\mathbf{z})]$ exists for all $l = 1,\dots,n$.
2. $\ell(\boldsymbol{\theta})$ is a well-defined function under these regularity conditions:
   R_1. $\ell(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0 \in \Theta$;
   R_2. Θ is compact;
   R_3. $\ell(\boldsymbol{\theta})$ is continuous;
   R_4. the empirical approximated log-likelihood function $\tilde{\ell}(\boldsymbol{\theta})$ converges uniformly in probability to $\ell(\boldsymbol{\theta})$.

It has to be noticed that, under concavity of the objective function $\tilde{\ell}(\boldsymbol{\theta})$, compactness ($R_2$) can be replaced by the assumption that

R_{2b}. the true parameter value $\boldsymbol{\theta}_0$ is an interior point of the parameter space, and the estimator $\hat{\boldsymbol{\theta}}$ is an interior point in a neighborhood containing $\boldsymbol{\theta}_0$ (see, e.g., Theorem 2.7 of Newey and McFadden [14]).

Let $\tilde{S}(\cdot)$ denote the approximated score vector according to the approximations (3.8); then we obtain

$$\begin{aligned}&\sum_{l=1}^{n} E_{\mathbf{z}|\mathbf{y}}\big[S_l(\hat{\boldsymbol{\theta}};\mathbf{z}_l)\big] = S(\hat{\boldsymbol{\theta}}) = \sum_{l=1}^{n}\big\{S_l(\hat{\boldsymbol{\theta}},\hat{\mathbf{z}}_l) + \cdots + \mathrm{O}\big(p^{-[k/3+1]}\big)\big\}\\ &\quad\Rightarrow\quad n^{-1}S(\hat{\boldsymbol{\theta}}) = n^{-1}\tilde{S}(\hat{\boldsymbol{\theta}}) + \mathrm{O}\big(p^{-[k/3+1]}\big)\end{aligned} \tag{B.1}$$

since $\hat{\boldsymbol{\theta}}$ is chosen such that $\sum_{l=1}^{n}\tilde{E}_{\mathbf{z}|\mathbf{y}}[S_l(\hat{\boldsymbol{\theta}};\mathbf{z}_l)] = \tilde{S}(\hat{\boldsymbol{\theta}}) = 0$. Under the regularity conditions in assumption 2 and provided that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathrm{o}_p(1)$, we can apply a Taylor series expansion in $S(\boldsymbol{\theta})$ around the true parameter vector $\boldsymbol{\theta}_0$:

$$S(\hat{\boldsymbol{\theta}}) = S(\boldsymbol{\theta}_0) + H\big(\boldsymbol{\theta}^*\big)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{B.2}$$

where $\boldsymbol{\theta}^*$ lies on the segment joining $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$, and

$$H(\boldsymbol{\theta}^*) = \frac{\partial S(\boldsymbol{\theta})}{\partial\theta}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \sum_{l=1}^{n}\frac{\partial S_l(\boldsymbol{\theta},\hat{\mathbf{z}}_l)}{\partial\theta}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \sum_{l=1}^{n} H_l\big(\boldsymbol{\theta}^*,\hat{\mathbf{z}}_l\big).$$

From equations (B.1) and (B.2), we obtain

$$\begin{aligned}&(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left\{n^{-1}\sum_{l=1}^{n} H_l\big(\boldsymbol{\theta}^*,\hat{\mathbf{z}}_l\big)\right\}^{-1}\big\{n^{-1}\big[S(\boldsymbol{\theta}_0) - S(\hat{\boldsymbol{\theta}})\big]\big\}\\ &\Rightarrow\quad (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left\{n^{-1}\sum_{l=1}^{n} H_l\big(\boldsymbol{\theta}^*,\hat{\mathbf{z}}_l\big)\right\}^{-1}\big[n^{-1}S(\boldsymbol{\theta}_0) + \mathrm{O}\big(p^{-[k/3+1]}\big)\big].\end{aligned}$$

In addition, under assumption 2, we have that, as $n \to \infty$, $n^{-1}H(\boldsymbol{\theta}^*) \to^p E_{\mathbf{y}}[H(\boldsymbol{\theta}_0)]$, where the expectation is taken with respect to $f(\mathbf{y};\boldsymbol{\theta})$, and $H(\boldsymbol{\theta}^*) = \sum_{l=1}^n H_l(\boldsymbol{\theta}^*, \hat{\mathbf{z}}_l)$. By further assuming that $E_{\mathbf{y}}\{H(\boldsymbol{\theta}_0)\}$ is non-singular, we obtain

$$\{n^{-1}H(\boldsymbol{\theta}^*)\}^{-1} \to^p E_{\mathbf{y}}\{H(\boldsymbol{\theta}_0)\}^{-1}.$$

It follows that

$$\begin{aligned}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= -E_{\mathbf{y}}[H(\boldsymbol{\theta}_0)]^{-1}[n^{-1}S(\boldsymbol{\theta}_0) + \mathrm{O}(p^{-[k/3+1]})] \\ &= \mathrm{O}_p[\max(n^{-1/2}, p^{-[k/3+1]})],\end{aligned}$$

where in the last step we use the fact that, under the regularity conditions 1, $n^{-1}S(\boldsymbol{\theta}_0) = \mathrm{O}_p(n^{-1/2})$, and $E_{\mathbf{y}}\{H(\boldsymbol{\theta}_0)\} = \mathrm{O}_p(1)$.

Appendix C: Development of the adaptive ML estimators for binary manifest variables

Let $\mathbf{y} = (y_1, \dots, y_p)^T$ be a vector of observed binary variables, having a Bernoulli distribution with expectation $\pi_j(\mathbf{z})$, $j = 1, \dots, p$. Using the canonical link function for Bernoulli distribution, we have

$$\pi_j(\mathbf{z}) = \frac{\exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z})}{1 + \exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z})}.$$

The scale parameter $\phi_j = 1$, such that the conditional distribution of each observed binary item given the latent variables $\mathbf{z}$ is

$$g_j(y_j|\mathbf{z};\boldsymbol{\theta}) = \exp[y_j(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z}) - \log(1 + \exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z}))], \qquad j = 1, \dots, p.$$

It follows that the approximated log-likelihood function (2.7) results

$$\begin{aligned}\tilde{\ell}(\boldsymbol{\theta}) = \sum_{l=1}^n \log\Bigg[& 2^{q/2}|\mathbf{T}_l| \\ & \times \sum_{t_1,\dots,t_q} \exp\Bigg(\sum_{i=1}^p y_{jl}(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\dots,t_q}) \\ & \qquad - \sum_{i=1}^p \log(1 + \exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\dots,t_q}))\Bigg) \\ & \times (2\pi)^{-q/2}\exp\left(-\frac{1}{2}\mathbf{z}^{*T}_{l,t_1,\dots,t_q}\mathbf{z}^*_{l,t_1,\dots,t_q}\right) w^*_{t_1}\cdots w^*_{t_q}\Bigg]\end{aligned} \tag{C.1}$$

$$
\begin{aligned}
&= \sum_{l=1}^{n} \Bigg\{ -\frac{q}{2} \log \pi \\
&\qquad + \log |\mathbf{T}_l| + \log \Bigg[\sum_{t_1,\dots,t_q} \exp\Bigg(\sum_{j=1}^{p} y_{jl}\big(\alpha_{0j} + \boldsymbol{\alpha}_j^T \mathbf{z}^*_{l,t_1,\dots,t_q}\big) \\
&\qquad\qquad - \sum_{j=1}^{p} \log\big(1 + \exp\big(\alpha_{0j} + \boldsymbol{\alpha}_j^T \mathbf{z}^*_{l,t_1,\dots,t_q}\big)\big) \\
&\qquad\qquad - \frac{1}{2} \mathbf{z}^{*T}_{l,t_1,\dots,t_q} \mathbf{z}^*_{l,t_1,\dots,t_q} \Bigg) w^*_{t_1} \cdots w^*_{t_q} \Bigg] \Bigg\},
\end{aligned}
$$

where the AGH nodes and weights are derived by the classical Gauss–Hermite nodes z_{t_k} and weights w_{t_k}, $k = 1, \dots, q$, as follows

$$
\mathbf{z}^*_{l,t_1,\dots,t_q} = \big(z^*_{l,t_1}, \dots, z^*_{l,t_q}\big)^T = \sqrt{2}\mathbf{T}_l (z_{t_1}, \dots, z_{t_q})^T + \hat{\mathbf{z}}_l
$$

and

$$
w^*_{t_k} = w_{t_k} \exp\big[z^2_{t_k}\big]
$$

with $\mathbf{T}_l$ derived by the Cholesky factorization of the matrix $\boldsymbol{\Psi}_l$, that is, $\boldsymbol{\Psi}_l = \mathbf{T}_l \mathbf{T}_l^T$. The modes $\hat{\mathbf{z}}_l$ are obtained for each subject through the iterative scheme

$$
\hat{\mathbf{z}}_l^{\mathrm{it}+1} = \hat{\mathbf{z}}_l^{\mathrm{it}} + \boldsymbol{\Psi}_l^{\mathrm{it}} L\big(\hat{\mathbf{z}}_l^{\mathrm{it}}\big),
$$

where “it” denotes the iteration counter,

$$
L\big(\hat{\mathbf{z}}_l^{\mathrm{it}}\big) = -\frac{\partial[\log g(\mathbf{y}_l|\mathbf{z}_l;\boldsymbol{\theta}) + \log h(\mathbf{z}_l)]}{\partial \mathbf{z}_l^T}\bigg|_{\mathbf{z}_l = \hat{\mathbf{z}}_l^{\mathrm{it}}} = -\sum_{j=1}^{p} \boldsymbol{\alpha}_j \left[y_{jl} - \frac{\exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_l^{\mathrm{it}})}{1 + \exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_l^{\mathrm{it}})} \right] + \hat{\mathbf{z}}_l^{\mathrm{it}}
$$

and

$$
\boldsymbol{\Psi}_l^{-1} = -\frac{\partial^2[\log g(\mathbf{y}_l|\mathbf{z}_l;\boldsymbol{\theta}) + \log h(\mathbf{z}_l)]}{\partial \mathbf{z}_l^T \partial \mathbf{z}_l}\bigg|_{\mathbf{z}_l = \hat{\mathbf{z}}_l^{\mathrm{it}}} = \sum_{j=1}^{p} \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^T \frac{\exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_l^{\mathrm{it}})}{1 + \exp(\alpha_{0j} + \boldsymbol{\alpha}_j^T \hat{\mathbf{z}}_l^{\mathrm{it}})} + \mathbf{I}.
$$

C.1. Regularity conditions for adaptive M-estimators in presence of binary data

Since the general theory of the M-estimators is here applied to a particular family of GLLVM, the regularity conditions on the log-likelihood function $\ell(\boldsymbol{\theta})$ given in Appendix B should be checked for the particular distribution of each observed variable (Huber, Ronchetti and Victoria-Feser [5]). For classical Laplace-based estimators, a formal proof of these conditions in presence of ordinal manifest variables is given by Huber, Scaillet and Victoria-Feser [6]. Following the main lines of that paper, we now prove how the empirical approximated log-likelihood (C.1) satisfies the regularity conditions for consistency and asymptotic normality of the corresponding M-estimators.

At this regard, we make use of the Lemma 2.2 by Newey and McFadden [14], according to which $\ell(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0 \in \Theta$ (condition R_1) if:

a_1. $\boldsymbol{\theta}_0$ is identified, that is, if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\boldsymbol{\theta} \in \Theta$, then $\ell(\boldsymbol{\theta}) \neq \ell(\boldsymbol{\theta}_0)$, and
a_2. $E[|\tilde{\ell}(\boldsymbol{\theta})|] < \infty$.

[a_1] Under our assumptions for the latent variables, $\boldsymbol{\theta}_0$ is identified.
[a_2] Let $\mathbf{z}^*$ be $\mathbf{z}^*_{l,t_1,\dots,t_q}$ and let $K(\mathbf{z}^*)$ denote $\log(1+\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*))$. We recall that $|\log(x)| \leq k(|x|+1)$ for a constant $k \geq 3$ and for any $x > 0$, and that $|\exp(x)| \leq \exp(|x|)$ for any $x \in \mathbb{R}$. Hence, based on (C.1),

$$\begin{aligned}
&\sum_{l=1}^{n}\left|\log\left[\sum_{t_1,\dots,t_q}\exp\left(\sum_{j=1}^{p} y_{jl}\left(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*\right)-\sum_{j=1}^{p}K\left(\mathbf{z}^*\right)-\frac{1}{2}\mathbf{z}^{*T}\mathbf{z}^*\right)w^*_{t_1}\cdots w^*_{t_q}\right]\right| \\
&\quad\leq \sum_{l=1}^{n}k\left(\sum_{t_1,\dots,t_q}\exp\left(\sum_{j=1}^{p}\left|y_{jl}\left(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*\right)\right|+\sum_{j=1}^{p}\left|K\left(\mathbf{z}^*\right)\right|+\frac{1}{2}\left\|\mathbf{z}^*\right\|\right)w^*_{t_1}\cdots w^*_{t_q}+1\right).
\end{aligned} \tag{C.2}$$

It can be noticed that $|K(\mathbf{z}^*)| \leq \log 2$ if $(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*) < 0$, and $|K(\mathbf{z}^*)| \leq |k_1||\boldsymbol{\alpha}_j|\|\mathbf{z}^*\| + |k_1||\alpha_{0j}| + |k_2| = |k_1||\boldsymbol{\alpha}_j|\|\mathbf{z}^*\| + \text{Const}$ if $(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*) > 0$ (using $\log(1+x) \leq k_1\log(x)+k_2$ for any constant $k_1 \geq \frac{1}{2}$ and $k_2 > 1$). Furthermore, $|y_{jl}(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*)| \leq |y_{jl}||\boldsymbol{\alpha}_j|\|\mathbf{z}^*\| + |y_{jl}||\alpha_{0j}| = |y_{jl}||\boldsymbol{\alpha}_j|\|\mathbf{z}^*\| + \text{Const}$. Using the definition of $\mathbf{z}^*$, we deduce that $E\|\mathbf{z}^*\| = \sum_{t_1,\dots,t_q}\mathbf{z}^{*T}_{l,t_1,\dots,t_q}\mathbf{z}^*_{l,t_1,\dots,t_q}w^*_{t_1}\cdots w^*_{t_q} < \infty$. Hence, based on log-normal moments, (C.2) is finite. Besides $|\log\det(\mathbf{T}_l)| < \text{Const}$, such that $E[|\tilde{\ell}(\boldsymbol{\theta})|] < \infty$.

Since the data are i.i.d. and Θ is compact (condition R_2), $\tilde{\ell}(\boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta}$ with probability one, and there is a function of the latent variables $d(\mathbf{z}^*)$ with $|\tilde{\ell}(\boldsymbol{\theta})| \leq d(\mathbf{z}^*)$ such that $E[d(\mathbf{z}^*)] < \infty$ (cf. proof of condition a_2). So, we deduce that $E[\tilde{\ell}(\boldsymbol{\theta})] = \ell(\boldsymbol{\theta})$ is continuous (condition R_3) and that $\tilde{\ell}(\boldsymbol{\theta})$ converges uniformly in probability to that quantity (condition R_4) (see Lemma 2.4 by Newey and McFadden [14]).

Asymptotic normality of the estimators imposes conditions on the Hessian of the empirical approximated log-likelihood function, that should be verified. Based on (C.1), by the computing the explicit expression of $\frac{\partial^2\tilde{\ell}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}$, it can be easily shown that there is a function $d(\mathbf{z}^*)$ with $|\frac{\partial^2\tilde{\ell}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}| < d(\mathbf{z}^*)$ such that $E[d(\mathbf{z}^*)] < \infty$ (as in the proof of condition a_2). As before, making use of Lemma 2.4 by Newey and McFadden [14], since the data are i.i.d. and Θ is compact, $\frac{\partial^2\tilde{\ell}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}$ is continuous at each $\boldsymbol{\theta}$ with probability one, and we can deduce that $E[\frac{\partial^2\tilde{\ell}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}]$ is continuous and the Hessian of the empirical approximated log-likelihood converges uniformly in probability to that quantity.

C.2. Score functions and second order Laplace estimators

In this specific case, the complete data score functions (2.8) are given by

$$S_l\big(\alpha_{0j};\mathbf{z}^*_{l,t_1,\dots,t_q}\big)=\left[y_{jl}-\frac{\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\dots,t_q})}{1+\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\dots,t_q})}\right],$$

$$S_l\big(\boldsymbol{\alpha}_j;\mathbf{z}^*_{l,t_1,\dots,t_q}\big)=\mathbf{z}^*_{l,t_1,\dots,t_q}\left[y_{jl}-\frac{\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\dots,t_q})}{1+\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\mathbf{z}^*_{l,t_1,\dots,t_q})}\right].$$

The corresponding score equations have not closed form solutions, and a quasi-Newton procedure is used to solve implicit equations.

In the simulation study, the performance of the adaptive-based estimators has been compared with second order Laplace estimators. According to (3.5), the latter have been derived by maximizing the following approximated log-likelihood function

$$\tilde{\ell}(\boldsymbol{\theta})=\sum_{l=1}^{n}\log\big\{(2\pi)^{q/2}|\boldsymbol{\Psi}_l^{1/2}\exp\big[-L(\hat{\mathbf{z}}_l)\big]\big[1+c_1p^{-1}+\mathrm{O}\big(p^{-2}\big)\big]\big\},$$

where the individual modes $\hat{\mathbf{z}}_l$ are obtained through the iterative scheme defined above, and

$$c_1=\sum_{m=2}^{3}\frac{(-1)^{m-1}}{(2m)!}L_{p_1}(\hat{\mathbf{z}}_l)\cdots L_{p_{m-1}}(\hat{\mathbf{z}}_l)L^{q_1}(\hat{\mathbf{z}}_l)\cdots L^{q_m}(\hat{\mathbf{z}}_l)$$

with $p_1|\cdots|p_{m-1}$ be a partition of $2m$ indices into $m-1$ blocks, each of size 3 or more, and $q_1|\cdots|q_m$ is a partition of $2m$ indices into m blocks, each of size 2. In particular, following the notation by Raudenbush, Yang and Yosef [19],

$$c_1=-\tfrac{1}{8}\operatorname{vec}^T[\boldsymbol{\Psi}_l\otimes\boldsymbol{\Psi}_l]\operatorname{vec}\big[L^{(4)}(\hat{\mathbf{z}}_l)\big]+\tfrac{5}{24}\operatorname{vec}^T[\boldsymbol{\Psi}_l\otimes\boldsymbol{\Psi}_l\otimes\boldsymbol{\Psi}_l]\operatorname{vec}\big[L^{(3)}(\hat{\mathbf{z}}_l)\otimes L^{(3)}(\hat{\mathbf{z}}_l)\big],$$

where

$$L^{(3)}(\hat{\mathbf{z}}_l)=-\sum_{j=1}^{p}\operatorname{vec}\big(\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T\big)\boldsymbol{\alpha}_j^T\frac{\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)[1-\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)]}{[1+\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)]^3}$$

and

$$\begin{aligned}L^{(4)}(\hat{\mathbf{z}}_l)=-\sum_{j=1}^{p}&\operatorname{vec}\big[\operatorname{vec}\big(\boldsymbol{\alpha}_j\boldsymbol{\alpha}_j^T\big)\boldsymbol{\alpha}_j^T\big]\\&\times\boldsymbol{\alpha}_j^T\frac{\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)[1-4\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)+\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)^2]}{[1+\exp(\alpha_{0j}+\boldsymbol{\alpha}_j^T\hat{\mathbf{z}}_l)]^4}.\end{aligned}$$

As for the adaptive-based estimators, the score equations of both the intercepts and factor loadings have not closed form solutions, and a quasi-Newton procedure has been used to solve implicit equations.

References

[1] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. *Monographs on Statistics and Applied Probability*. London: Chapman & Hall. MR1010226

[2] Bianconcini, S. and Cagnone, S. (2012). Estimation of generalized linear latent variable models via fully exponential Laplace approximation. *J. Multivariate Anal.* **112** 183–193. MR2957295

[3] de Bruijn, N.G. (1981). *Asymptotic Methods in Analysis*, 3rd ed. New York: Dover. MR0671583

[4] Evangelou, E., Zhu, Z. and Smith, R.L. (2011). Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation. *J. Statist. Plann. Inference* **141** 3564–3577. MR2817363

[5] Huber, P., Ronchetti, E. and Victoria-Feser, M.P. (2004). Estimation of generalized linear latent variable models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 893–908. MR2102471

[6] Huber, P., Scaillet, O. and Victoria-Feser, M.P. (2009). Assessing multivariate predictors of financial market movements: A latent factor framework for ordinal data. *Ann. Appl. Stat.* **3** 249–271. MR2668707

[7] Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Comput. Statist. Data Anal.* **52** 5066–5074. MR2526575

[8] Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: An example. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 325–335. MR1856328

[9] Liu, Q. and Pierce, D.A. (1994). A note on Gauss–Hermite quadrature. *Biometrika* **81** 624–629. MR1311107

[10] Martins, T.G., Simpson, D., Lindgren, F. and Rue, H. (2013). Bayesian computing with INLA: New features. *Comput. Statist. Data Anal.* **67** 68–83.

[11] Moustaki, I. and Knott, M. (2000). Generalized latent trait models. *Psychometrika* **65** 391–411. MR1792703

[12] Muthén, L.K. and Muthén, B.O. (2010). *Mplus User's Guide*, 6th ed. Los Angeles, CA: Muthén & Muthén.

[13] Naylor, J.C. and Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distributions. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **31** 214–225. MR0694917

[14] Newey, W.K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, *Vol. IV* (R.F. Engle and D. McFadden, eds.). *Handbooks in Econom.* **2** 2111–2245. Amsterdam: North-Holland. MR1315971

[15] Pinhero, J.C. and Bates, D.M. (1995). Approximation to the log-likelihood function in the nonlinear mixed effects model. *J. Comput. Graph. Statist.* **4** 12–35.

[16] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

[17] Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata*, 3rd ed. College Station, TX: Stata Press.

[18] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *J. Econometrics* **128** 301–323. MR2189555

[19] Raudenbush, S.W., Yang, M.L. and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *J. Comput. Graph. Statist.* **9** 141–157. MR1826278

[20] Rizopoulos, D., Verbeke, G. and Lesaffre, E. (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 637–654. MR2749911

[21] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Boca Raton, FL: Chapman & Hall/CRC. MR2130347

[22] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 319–392. MR2649602

[23] Schilling, S. and Bock, R.D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika* **70** 533–555. MR2272503

[24] Shun, Z. (1997). Another look at the salamander mating data: A modified Laplace approximation approach. *J. Amer. Statist. Assoc.* **92** 341–349.

[25] Shun, Z. and McCullagh, P. (1995). Laplace approximation of high-dimensional integrals. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57** 749–760. MR1354079

[26] Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Interdisciplinary Statistics*. Boca Raton, FL: Chapman & Hall/CRC. MR2059021

[27] Tauchen, G. and Hussey, R. (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica* **59** 371–396. MR1097533

[28] Tierney, L., Kass, R.E. and Kadane, J.B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* **84** 710–716. MR1132586

[29] Vonesh, E.F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* **83** 447–452. MR1439795

Received October 2012 and revised May 2013