Risk management papers

A generalised latent Poisson factor modelling approach for default correlations in credit portfolios

Mohamed Saidane

Associate Professor, College of Business and Economics, Qassim University, Kingdom of Saudi Arabia

Dr Mohamed Saidane received his PhD degree in applied mathematics from the Languedoc University of Science and Technology in Montpellier, France in 2006. Currently, he is an associate professor in the Department of Management Information Systems and Production Management at Qassim University, Saudi Arabia and is a full professor of applied statistics at the University of Tunis, Tunisia. His research interests include model-based clustering and classification, statistical learning from complex data, latent structure modelling, with applications to business analytics. His publications have appeared in *Applied Stochastic Models in Business and Industry*, *Advances in Statistical Analysis*, *Communications in Statistics*, *Computational Economics* and other academic journals.

Department of Management Information Systems and Production Management, College of Business and Economics, Qassim University, P.O. Box 6666, Buraidah, 51452, Kingdom of Saudi Arabia
E-mail: m.saidane@qu.edu.sa

Abstract   Default risk is one of the major concerns for lending institutions and banking regulators. This paper focuses on the analysis of default data, using a new approach based on generalised latent Poisson factor models. In this case, the correlation structure of the default events is driven by a small number of common latent factors. Conditional to these factors, the defaults become independent and each default sequence is fitted to a generalised linear model with Poisson response and log-link function. This model provides a flexible framework for the computation of the value-at-risk and the expected shortfall of a credit portfolio. The practical implementation of the proposed local Fisher scoring estimation algorithm is illustrated by a Monte Carlo simulation study. Then, a real scenario, with default data taken from a large database provided by Standard & Poor's, is used to analyse the empirical behaviours of the different risk measures. The achieved results show promising performance.

Keywords: *default correlation, factor analysis, generalised linear models, expectation-maximisation algorithm, credit value-at-risk, expected shortfall*

INTRODUCTION

The correlation between default events is an essential factor in the assessment of portfolio credit risk. During the last two decades, many research works have been devoted to modelling this complex phenomenon. An excellent literature review in this area can be found in Nguyen and Zhou[1] and Jakob,[2] where the available models are classified into two categories: mixture models and latent structure models. Moody's–KMV model,[3] also described by Frey and McNeil,[4] is a variant of the latent variable model proposed by Merton.[5] These models assume that the default probabilities of the different obligors are driven by a series of common economic factors. Given these common factors, defaults are assumed to be conditionally independent. Hence, it can be argued here that the correlations are directly explained by the dependence on the common economic factors. A standard example is the 'large homogeneous portfolio' model developed by Vasicek.[6]

A number of other default correlation models have been proposed in the literature, such as those of

McNeil *et al.*[7] and Bluhm *et al.*[8] and their references. Works by Gupton *et al.*[9] and the KMV model of Crosbie[10] including the Bernoulli model developed by CreditMetrics, the 'Poisson mixture model' proposed by the Credit Suisse Financial Products[11] and the works of Das *et al.*[12] and Duffie *et al.*[13] may also be mentioned.

Copula models for credit risk measurement, were introduced for the first time by Li.[14] In particular, he developed a generalised defaultable bond pricing formula based on a Gaussian copula framework with exponentially distributed default times. This approach was extensively used on Wall Street until the beginning of 2008. After the sub-prime mortgage crisis, an extensive literature has been developed around the use of dynamic copula models. For a detailed and rich literature review on this topic, interested readers can refer to Muhajir *et al.*[15]

Most of the work presented in the previous literature focuses on modelling the default correlations in a cross-sectional framework within a specific sector. However, there is a minor literature on modelling the default correlations between several sectors: the case where the default of a bond in a specified sector affects the default probabilities of certain other bonds in the other sectors. Thus, modelling the impact of defaulted bonds on the default probabilities of other bonds is certainly an interesting topic. This is particularly the case during a default crisis where this impact becomes more marked and the defaulting bonds become more likely to trigger the defaults of the surviving ones.

This paper proposes a general framework to model default correlations in bond portfolios, estimate the number of defaults and price the risks of default. The number of defaulted bonds in each sector are assumed to be produced from a general process that involves observed and non-observed data following an exponential family distribution. In this particular case, the Poisson distribution is used to model the number of default events occurring within a given period of time (assessment period) and the common latent factors are assumed to be independently distributed, following a multivariate normal distribution with zero mean and identity covariance matrix. To price the default risk, two principal risk measures will be used, namely, value-at-risk (VaR) and expected shortfall (ES). This approach is general enough and can be applied to large portfolios consisting of many bonds, with reasonable computational time and effort.

The rest of this paper is organised as follows. A detailed description of the generalised latent Poisson factor model and its main characteristics will be presented. The estimation procedure, using a local approximation of the expectation maximisation (EM) algorithm, will be discussed. Thereafter, a Monte Carlo simulation strategy for the risk measures will be derived. The following section will be devoted to a set of extensive numerical experiments designed to evaluate the performance of the proposed EM algorithm. Moreover, the VaR and ES results given by this approach will be also discussed, using the Standard & Poor's CreditPro 6.2 database. Finally, some concluding remarks and suggestions on future research directions will be made.

MATHEMATICAL FORMULATION OF THE PROPOSED MODEL

Throughout this paper, the number of defaults in a loan portfolio is modelled by first forming homogeneous groups of bonds according to their economic sector. Within each group, individual defaults are assumed to be independent, following a Bernoulli distribution.

For each time $t = 1, \ldots, T$, this framework assumes that the total number of defaults in the different q sectors, $\gamma_t = (\gamma_{it})_{i=1,q}$, are generated by fewer $k < q$ latent factors $f_t = (f_{jt})_{j=1,k}$. Using these factors as predictors, each sequence γ_i is fitted by a generalised linear model (GLM) with Poisson response and log-link function. In order to avoid identification problems, common factors are assumed to follow independent Gaussian distributions with zero mean and unit variance:

$$\forall t, \; f_t \sim N(0, I_k) \tag{1}$$

Conditional to these factors, γ_{it} are independently distributed according to an exponential family with density function:

$$l_i\left(\gamma_{it} \mid \delta_{it}, \phi_i\right) = exp\left\{\frac{\gamma_{it}\delta_{it} - b_i\left(\delta_{it}\right)}{\phi_i} + c_i\left(\gamma_{it}, \phi_i\right)\right\} \tag{2}$$

where,

$$\mu_{it} = E(y_{it}) = b_i'(\delta_{it}) \tag{3}$$

$$Var(y_{it}) = \phi_i b_i''(\delta_{it}) = \phi_i b_i'' \left[b_i'^{-1}(\mu_{it}) \right] \tag{4}$$

In this case, δ_{it} and ϕ_i denote, respectively, the canonical and the dispersion parameters, and $b_i(\cdot)$ and $c_i(\cdot)$ are specific functions characterising the probability distribution of the variable y_{it} in the exponential family (see eg Fahrmeir and Tutz[16]).

Given the conditional independence of $(y_{it})_{i=1,q}$ with respect to the common latent factors f_t, the conditional covariance matrix of y_t can be obtained as follows:

$$Var(y_t \mid f_t) = diag\left\{\phi_i v_i(\mu_{it})\right\}_{i=1,q} \tag{5}$$

where, $v_i(\mu_{it}) = b_i''(b_i'^{-1}(\mu_{it}))$. In the case of a Poisson distribution, the expression of μ_{it} is given by $\exp(\delta_{it})$, $v_i(\mu_{it}) = \mu_{it}$ and $\phi_i = 1$ (for more details, see McCullagh and Nelder,[17] Chapter 10).

N.B. In this paper, the operators $'$ and $''$ denote, respectively, the first and second order derivatives of $b_i(x)$ and $g(x)$ with respect to x. However, in matrix notation A' denotes the transpose of A.

Linear predictors

In this framework, the different default sequences y_i are assumed to be driven by specific predictors η_i derived as linear combinations of some common latent factors:

$$\forall i,t: \eta_{it} = \gamma_i + \lambda_i' f_t; \; \lambda_i \in \mathbb{R}^k \tag{6}$$

In the previous literature, various approaches have been developed to identify the common factors. Most of these approaches have used principle component analysis and some predefined weights from the economic theory to extract the underlying risk factors from a set of macroeconomic variables, such as the inflation rate, business cycle patterns, etc. This paper presents a solution based on the EM algorithm and a backtesting procedure.

A more general specification of these predictors can also be obtained by including various other covariates, but for simplicity they have been taken as constants in this paper. In the following, the specific

mean vector will be denoted by $\gamma = (\gamma_1, \ldots, \gamma_q)'$, the $(q \times T)$ common factor matrix by $\mathcal{F} = (f_1, \ldots, f_T)$, the $(q \times k)$ factor loading matrix by $\Lambda = (\lambda_1, \ldots, \lambda_q)'$, the q-vector containing the predictor variables at time t by $\eta_t = (\eta_{it})_i$, and the $(q \times T)$ predictors' matrix by $\eta = (\eta_{it})_{i,t} = (\eta_1, \ldots, \eta_T)$. In this case, equation (6) can be expressed in temporal form as follows:

$$\forall t, \; \eta_t = \gamma + \Lambda f_t \tag{7}$$

and in global form as:

$$\underset{(q,T)}{\eta} = \underset{(q,1)}{\gamma} \underset{(1,T)}{1_T'} + \underset{(q,k)}{\Lambda} \underset{(k,T)}{\mathcal{F}} \tag{8}$$

Link function

The mean of the response variable y_{it} is linked to the set of common latent factors by a linear predictor η_{it} with unknown parameters and a specific link function g_i:

$$\forall i,t \quad \eta_{it} = g_i(\mu_{it}) \tag{9}$$

In the case of $\eta_i = \delta_i$, the function $g_i(\cdot)$ is called canonical link function. Moreover, it can be shown from equations (3) and (9) that the linear predictor can be explicitly obtained by:

$$\eta_{it} = g_i(b_i'(\delta_{it})) \tag{10}$$

where the canonical link function is $g_i = b_i'^{-1}$. In the Poisson GLM, the mean of the response variable is linked to the linear predictor through the natural logarithm, ie $g(x) = \log(x)$.

ESTIMATION IN GENERALISED LINEAR LATENT FACTOR MODELS

The parameters of standard and conditionally heteroskedastic latent factor models (FMs) are estimated using a variety of maximum likelihood inference methods based on the EM principle (see eg Saidane and Lavergne[18–20] and Saidane[21,22]). GLMs are also estimated using maximum likelihood methods based on the Fisher's score algorithm (McCullagh and Nelder[23]). However, combining factorial models and GLMs leads to more complex

models and estimation algorithms, such as those proposed by Wedel and Kamakura[24]; Genton and Ronchetti[25] and Moustaki and Victoria-Feser.[26] All these methods consider only single factor models and they are computationally complex and require high computing power and long execution time, which precludes their use for multifactor models and big data analysis.

Inspired by the previous literature, this paper proposes a quicker estimation procedure based on a local approximation of the iterative EM algorithm developed by Dempster *et al.*[27] In a first step, the expected values of the common latent factors will be used as predictors and the parameters in the conditional GLM will be estimated using maximum likelihood estimators. Then, in a second step, the latent random state of the common factors will be restored and the parameter estimates of the linear predictor will be updated using the EM algorithm. These steps will be repeated until convergence is reached.

Fisher scoring algorithm for the GLM

To begin with, one can consider the simple univariate GLM for given data (y_t, x_t). In this case, the predictors $x_t = (x_{1t}, \ldots, x_{kt})'$ are assumed to be observed without error. Let $X = (x_1, \ldots, x_t, \ldots, x_T)'$; $\mu_t = \mathbb{E}(y_t)$ and the linear predictor $\eta = X\beta$, $\beta \in \mathbb{R}^k$. At each time t, the linear predictor η_t is related to the mean μ_t through a specified link function $\eta_t = g(\mu_t)$ which can also be expressed as $x_t'\beta = g(b'(\delta_t))$. The log-likelihood function of this model is explicitly given by:

$$\mathcal{L} = \sum_{t=1}^{T} \mathcal{L}_t(y_t \mid \delta_t, \phi) = \sum_{t=1}^{T} \left[\frac{y_t \delta_t - b(\delta_t)}{\phi} + c(y_t, \phi) \right] \quad (11)$$

and the first derivatives with respect to β_j can be obtained using the following rule:

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{t=1}^{T} x_{tj} \frac{1}{g'(\mu_t)^2 Var(y_t)} g'(\mu_t)(y_t - \mu_t)$$

In the special case of a Poisson distribution, if one denotes by:

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

and

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

Then, it holds by construction that:

$$W_\beta = diag\left[\frac{1}{exp(x_t'\beta)} \right]_{t=1,T} \quad (12)$$

and the likelihood equations can be solved as follows:

$$\nabla_\beta \mathcal{L} = 0 \Leftrightarrow X'W_\beta^{-1} \frac{\partial \eta}{\partial \mu}(y - \mu) = 0 \quad (13)$$

Since this equation system depends nonlinearly on β it cannot be solved directly and an iterative method, such as the Fisher scoring algorithm (FSA), must be implemented:

$$\beta^{[e+1]} = \beta^{[e]} - \left(E\left(\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta'} \right)^{[e]} \right)^{-1} \left(\frac{\partial \mathcal{L}}{\partial \beta} \right)^{[e]}$$

$$= \left(X'W_{\beta^{[e]}}^{-1}X \right)^{-1} X'W_{\beta^{[e]}}^{-1} z^{[e]}$$

where $\beta^{[e]}$ denotes the estimated parameter vector at the e-th iteration and

$$z^{[e]} = X\beta^{[e]} + \left(\frac{\partial \eta}{\partial \mu} \right)^{[e]} (y - \mu^{[e]}) \quad (14)$$

In this case, equation (13) becomes:

$$X'W_\beta^{-1}(z_\beta - X\beta) = 0 \quad (15)$$

Given z_β, the equation system (15) may be viewed as a normal generalised least squares (GLS) equation system of the linear model $\mathcal{M}$:

$$z_\beta = X\beta + \zeta$$

with $E(\zeta) = 0$ and $Var(\zeta) = W_\beta$, since $Var(\zeta_t) = Var(z_{\beta,t}) = g'(\mu_t)^2 Var(y_t)$. The key step in each iteration of the FSA is to solve with respect to β the system $X'W_{\beta^{[e]}}^{-1}(z_{\beta^{[e]}} - X\beta) = 0$, and to update thereafter the value of β in W_β and z_β with the optimal solution. In the sequel, $\mathcal{M}^{[e]}$ will denote the linearised model at the current iteration [e]: $z_{\beta^{[e]}} = X\beta + \zeta^{[e]}$ with $E(\zeta^{[e]}) = 0$ and $Var(\zeta^{[e]}) = W_{\beta^{[e]}}$.

In the multivariate case, one can refer to the works by Niku *et al.*[28] and Huber *et al.*[29] where the response variables $\{y_1,\ldots,y_q\}$ are assumed to be (conditionally) independent given the covariates $\{x_1,\ldots,x_k\}$. From this assumption, it follows that:

$$l(y_t|\eta_t) = \prod_{i=1}^{q} l_i(y_{it}|\eta_{it}); \quad \forall t = 1,T \qquad (16)$$

and

$$l(y|\eta) = \prod_{i=1}^{q} l_i(y_i|\eta_i)$$

As a consequence, the corresponding linearised model $\mathcal{M}$ can be expressed as follows:

$$z_{i\beta} = X\beta_i + \zeta_i$$

where $\forall i = 1,q : E(\zeta_i) = 0$ and $Var(\zeta_i) = W_{i\beta}$ with $W_{i\beta} = diag\left[g_i'(\mu_{it})^2 Var(y_{it})\right]_{t=1,T} = diag\left[g_i'(\mu_{it})^2 \phi_i v_i(\mu_{it})\right]_{t=1,T}$

and the FSA can be used for the estimation of the model parameters.

For the initialisation of β, one can use the ordinary least squares estimates of the regression of $g_i(y_i)$ on X. In this case, the first order Taylor development of $g_i(y_i)$ around μ_i writes: $g_i(y_i) = g_i(\mu_i) + g_i'(\mu_i)(y_i - \mu_i) = z_i$. Thus, in order to avoid the indetermination of g_i owing to the presence of zero-responses among the data, in the case of a log-link function, one can take: $z_{it}^{[0]} = g_i(ay_{it} + (1-a)\overline{y}_i)$ for all t.

Estimation of the Gaussian FM

Following Saidane[30] (Chapter 2), the general structure of the standard latent factor analysis model can be summarised as follows:

$$y_t = \gamma + \Lambda f_t + \varepsilon_t$$

and

$$\begin{pmatrix} y_t \\ f_t \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma \\ 0 \end{pmatrix}; \begin{pmatrix} \Lambda\Lambda' + \Psi & \Lambda \\ \Lambda' & I_k \end{pmatrix} \right)$$

where, $\forall t = 1,T$, $Var(\varepsilon_t|f_t) = \Psi$. The estimation of the specific mean vector γ, the factor loading matrix Λ and the specific covariance matrix Ψ can be carried out by solving the first derivatives set of the conditional expectation of the complete data log-likelihood function as follows:

$$\nabla_{\gamma,\Lambda,\Psi}\left[\sum_{t=1}^{T} E(\log l(y_t,f_t)|y_t) \right] = 0 \qquad (17)$$

In the case of an heteroskedastic Gaussian FM, this function is given by:

$$\log l(y_t,f_t) \approx \omega - \frac{1}{2}\sum_{t=1}^{T}\left[\log|\Psi_t| + (y_t - \Lambda f_t - \gamma)' \Psi_t^{-1}(y_t - \Lambda f_t - \gamma) \right]$$

where $\forall t = 1,T$, $\Psi_t = Var(\varepsilon_t|f_t)$ and the component ω does not depend on the parameters γ, Λ and Ψ_t. Following Saidane[31] and Saidane and Lavergne,[32] when the first derivative of equation (17), with respect to γ, is equated to zero, one can find the updated estimate of the specific mean:

$$\hat{\gamma}^{[e+1]} = \left[\sum_{t=1}^{T}\Psi_t^{[e]-1} \right]^{-1} \sum_{t=1}^{T}\Psi_t^{-1}\left(y_t - \hat{\Lambda}^{[e]}\tilde{f}_t^{[e]}\right)$$

Similarly, equating to zero the first derivative with respect to Λ gives:

$$\sum_{t=1}^{T}\Psi_t^{-1}\left[(y_t - \hat{\gamma})\tilde{f}_t' \right] = \sum_{t=1}^{T}\Psi_t^{-1}\hat{\Lambda}\tilde{S}_t$$

The row by row procedure proposed by Saidane[33,34] can be used to estimate the factor loading matrix Λ, when the number of common factors exceeds 1. In the case of 2 factors, for example, given the fact that $\Psi_t^{-1} = diag(\Psi_t^{-1})$, the *i*-th row of the above matrix can be written as follows:

$$\sum_{t=1}^{T}\frac{y_{it} - \gamma_i^{[e]}}{\Psi_t^{i[e]}}\left(\tilde{f}_t^{1[e]} \quad \tilde{f}_t^{2[e]} \right) = \left(\hat{\lambda}_{i1}^{[e+1]} \quad \hat{\lambda}_{i2}^{[e+1]} \right)\sum_{t=1}^{T}\frac{1}{\Psi_t^{i[e]}}\begin{bmatrix} \tilde{S}_t^{11[e]} & \tilde{S}_t^{12[e]} \\ \tilde{S}_t^{21[e]} & \tilde{S}_t^{22[e]} \end{bmatrix}$$

where $\hat{\Lambda}^{[e]} = \left(\left(\hat{\lambda}_{ij}^{[e]} \right) \right)_{i,j}$ and $\tilde{S}_t^{[e]} = \left(\left(\tilde{S}_t^{ij[e]} \right) \right)_{i,j}$. The resolution of this equation system gives the i-th row of the updated factor loading matrix $\hat{\Lambda}^{[e+1]}$.

Finally, to obtain the updated specific covariance matrix, the above log-likelihood function is differentiated with respect to Ψ_t^{-1} and equated to zero. Solving for Ψ_t yields:

$$\hat{\Psi}_t^{[e+1]} = diag \left[\gamma_t \gamma_t' - \begin{bmatrix} \Lambda'^{[e]} & \hat{\gamma}^{[e]} \end{bmatrix} \begin{bmatrix} \gamma_t \tilde{f}_t^{[e]\prime} & \gamma_t \end{bmatrix}' \right]$$

Estimation of a generalised latent factor model

Application of the EM algorithm to the standard FM does not pose any computational difficulties since the required conditional expectations can be performed analytically in this case. One can also take advantage from the normality assumption, which may be used within each iteration to estimate a linearised GLM via the GLS algorithm discussed in section 'Fisher scoring algorithm for the GLM'. The proposed algorithm alternates between two main steps.

The linearisation step

Given the actual values of γ, Λ, F, the working variable z can be obtained from equation (14) as follows:

$$\forall i = 1, q : z_{i,F} = \gamma_i 1_T' + \lambda_i' F + g'(\mu_{i,F})(\gamma_i - \mu_{i,F})$$

Let $\zeta_{i,F} = g'(\mu_{i,F}) \epsilon_{i,F}$, where $\epsilon_{i,F} = \gamma_i - \mu_{i,F}$; $\forall t, z_t = (z_{1t}, \ldots, z_{qt})$ and $z = (z_1, \ldots, z_t, \ldots, z_T)'$. Given z, F and $Var(\zeta \mid F)$, the linearised form of the model is obtained as follows:

$$\forall t = 1, T : z_t = \gamma + \Lambda f_t + \zeta_t \tag{18}$$

with the conditional covariance matrix given by:

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

where, $Var(\zeta_{iF} \mid F) = W_{i,\Lambda F}$ and in the case of Poisson outcomes and a log-link function, the matrix W is specified as indicated in equation (12).

The estimation step

In this step, given the values of z and $Var(\zeta)$, the latent random state of the common factors will be restored:

$$\forall t, f_t \sim N(0, I_k)$$

and the model (18) can be written as an heteroskedastic FM, with mean $E(z_t) = \gamma$ and a covariance structure dependent on the observed data: $\Psi_t = g'(\mu_{i,f_t})^2 Var(\epsilon_{it} \mid f_t)$. In the case of a canonical link and a Poisson distribution, it can be shown that:

$$Var(\epsilon_{it} \mid f_t) = Var(\gamma_{it} \mid f_t) = b_i''\left(b_i'^{-1}(\mu_{it}) \right) = \frac{1}{g_i'(\mu_{it})}$$

which implies that $\Psi_t = diag\left(g'(\mu_{it})^2 \right)$.

THE LOCAL EM ALGORITHM

i. Parameters' initialisation:

Compute: $\forall t = 1, T$; $\forall i = 1, q : z_{it}^{[0]} = g(\gamma_{it})$, and for instance: $\forall t = 1, T : \Psi_t^{[0]} = I_q$.

In the case of a canonical link function g, the specific covariance matrix may be correctly initialised using: $\forall t = 1, T, \Psi_t^{[0]} = diag\left(g_i'(\gamma_{it})^2 v_i(\gamma_{it}) \right)_{i=1,q}$.

ii. Given the values of z and $Var(\zeta)$, the latent random state of the common factors will be restored. Then, an EM step will be taken in order to estimate F from the resulting linearised heteroskedastic FM: $z_t = \gamma + \Lambda f_t + \zeta_t$, with $Var(\zeta_t) = \Psi_t$, $\forall t = 1, T$.

iii. After a conditional GLM is obtained using the value of F, computed in the previous step, one can implement the FSA in order to update the specific mean vector γ and the factor loading matrix Λ, using the conditional covariance matrix $Var(z_t \mid F_t) = Var(\zeta_t) = \Psi_t$.

iv. Once updated estimates for γ, Λ, F are made, one computes z and $Var(\zeta)$:

$$\forall i = 1, q : \epsilon_{i,F} = \gamma_i - \mu_{i,F} \; ; \; \zeta_{i,F} = g'(\mu_{i,F}) \epsilon_{i,F} \; ; \; z_{i,F}$$
$$= \gamma_i 1_T' + \lambda' F + \zeta_{i,F}$$

$$\forall t = 1, T : Var(\zeta_t) = \Psi_t = diag\left(g'(\mu_{i,f_t}) \right)_{i=1,q}$$

Go to (ii).

MONTE CARLO VaR AND ES COMPUTATIONS FOR CREDIT PORTFOLIOS

This section follows the methodology outlined in Saidane[35] to compute the VaR and ES of a credit portfolio. In order to evaluate these measures, each bond in the loan portfolio is assumed to carry one dollar. To achieve a balance between efficiency and parsimony, two simulation methods will be implemented and compared in a rolling window design with $n = 50.000$ random default scenarios: the classical Monte Carlo method (CMC) and the proposed Monte Carlo-based generalised latent Poisson factor modelling (GLPFM) approach.

The classical Monte Carlo approach

The steps of the CMC approach are detailed as follows:

1. Define the VaR significance level α.
2. Simulate the number of defaults from time t to time $t+1$ by generating random variates from the Poisson distribution defined by λ (the empirical default rate calculated as an average on a rolling window basis). The following recursive rules can be used to get successive probabilities from the Poisson distribution:

$$p_{i+1} = p(X = i+1) = \frac{\lambda}{i+1} p_i \; ; \qquad i \geq 0$$

This leads to the following method:

(i) Firstly, a uniform pseudo random number U is generated.
(ii) Then, initial values for $i = 0$, $p_0 = e^{-\lambda}$ and $R_0 = p_0$, should be provided.
(iii) If $U \leq R_i$, then deliver $X = R_i$. Return to step (i).
(iv) Else, the loop variable 'i' is incremented by 1: $p_{i+1} = \lambda p_i / (i+1)$ and $R_{i+1} = R_i + p_{i+1}$.

Return to step (iii).

3. After that, the simulated portfolio losses at time $t+1$, can be computed as follows:

$$L_{t+1,s}^p = L_{1t+1}^s + L_{2t+1}^s + \ldots + L_{qt+1}^s$$

where L_{it+1}^s, $(s = 1,\ldots,n)$, denote the simulated number of defaulted bonds in each sector $i = 1,\ldots,q$ for the period $t+1$.

4. Finally, the simulations should be arranged in an increasing order and the α per cent worst (highest) losses $L_{t+1,s}^p$ must be excluded. Therefore, the predicted VaR for the period $t+1$, will be computed as the maximum of the remaining losses. The ES is calculated by averaging the α per cent worst portfolio losses $L_{t+1,s}^p$.

Monte Carlo-based GLPFM approach

The steps of the proposed GLPFM-based approach are detailed as follows:

1. Define the VaR significance level α.
2. Generate from the distribution of the common factors $f_t^s \sim \mathcal{N}(0, I_k)$, n different random scenarios using the optimal specification at time t obtained by the BIC criterion (Schwarz[36]), evaluated after each EM cycle.
3. Simulate n different random scenarios for the linear predictor η_i^s, using the latest parameter estimates given by the EM algorithm at time t:

$$\eta_{i+1}^s = \gamma + \Lambda f_{i+1}^s$$

Then, n different values for the means of the portfolio's components can be computed as follows:

$$\mu_{it+1}^s = exp(\eta_{it+1}^s) \qquad \forall i = 1,\ldots,q$$

4. Generate from the Poisson distribution the default sequences for the q portfolio's sectors:

$$\gamma_{it+1}^s = rpois(1, \mu_{it+1}^s)$$

$$\forall s = 1,\ldots,n, \; \forall i = 1,\ldots,q$$

In this case, the different portfolio losses at time $t+1$, can be computed as follows:

$$L_{t+1,s}^p = \gamma_{1t+1}^s + \gamma_{2t+1}^s + \ldots + \gamma_{qt+1}^s$$

where γ_{1t+1}^s, γ_{2t+1}^s, $\ldots$, γ_{qt+1}^s $(s = 1,\ldots,n)$, denote the simulated number of defaulted bonds in the q sectors for the period $t+1$.

5. Finally, the simulations should be arranged in an increasing order and the α per cent worst

(highest) losses $L^p_{t+1,s}$ must be excluded. Therefore, the predicted VaR for the period $t+1$, will be computed as the maximum of the remaining losses. The ES is calculated by averaging the α per cent worst portfolio losses $L^p_{t+1,s}$.

NUMERICAL EXPERIMENTS

The major contribution of this paper consists in the development of a new Monte Carlo-based GLPFM framework for the estimation of the VaR and ES of a credit portfolio. To investigate the potential utility of the estimation algorithm and the effectiveness of the proposed risk assessment approach, synthetic and real data experiments were conducted. All the computational results in this paper were obtained using the R software (Version 4.2.2).

Synthetic data

In order to evaluate the accuracy and stability of the estimates obtained via the proposed local EM algorithm, a simulation study has been developed. For this purpose, a GLPFM is used to generate correlated Poisson outcomes from different factorial specifications with $k = 1, 2, 3$, $q = 8$ and $n = 200$ observations. To stop the EM iterations, the following convergence criterion is applied:

$$\Delta = \max_{j \in \{1, \ldots, q\}} \left\{ \sum_{t=1}^{n} \left(f^{[e+1]}_{jt} - f^{[e]}_{jt} \right)^2 \right\}$$

The algorithm stops when the value of Δ, between two successive iterations, becomes less than 10^{-5}. For the initialisation of the working variable z, and in order to avoid the indetermination of the log-link function owing to the presence of zero-responses among the data, the following guess values were used:

$$\forall i = 1, q; \quad t = 1, T \qquad z^{[0]}_{it} = \log \left[a \gamma_{it} + (1-a) \overline{y}_i \right]$$

where the coefficient is $a \in [0,1]$. For the initialisation of the EM algorithm, random perturbation of the simulation parameters were used as initial guess values. In order to identify the optimal number of common latent factors, which can be used to better fit the data, various configurations of the GLPFM were estimated using the local EM algorithm (with $k = 1, 2, 3$), on the different training datasets. To analyse the results of these experiments, empirical correlations between the estimated and true parameters were computed at the end of each EM cycle.

From the results given in Table 1, it can be seen that these correlations are strongly affected by the model parametrisation: correlations obtained when using the correct specification are larger than those given by the other specifications. Moreover, it can be seen also that the estimation error becomes significantly smaller when the accurate specification is used.

Figure 1, reports the results from the regressions of the simulated factors f_t on the estimated ones $f_t^{(e)}$:

Table 1: The estimation errors and the correlations of the true parameters with the estimated ones obtained from a GLPFM with one, two and three common factors

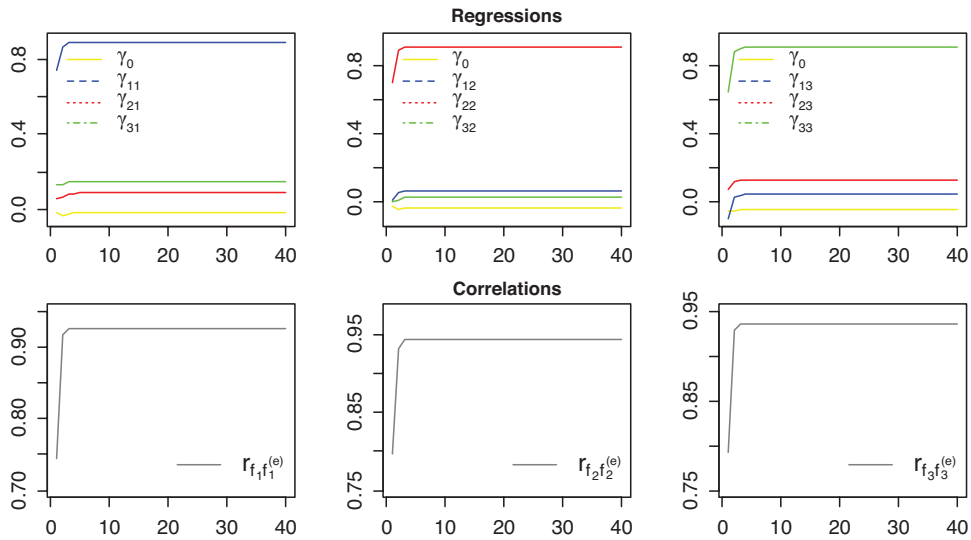| True model	Estimated model	Errors			Correlations		
		γ	Λ	γ	Λ		
3	3	0.0085	0.3240	0.9898	0.9984	0.9936	0.9978
	2	0.0255	1.2914	0.9536	0.9776	0.8373	
	1	0.0823	1.8412	0.9066	0.6735		
2	3	0.0078	0.0115	0.9829	0.9880	0.9904	
	2	0.0068	0.0095	0.9830	0.9884	0.9954	
	1	0.0073	0.3238	0.9696	0.6857		
1	3	0.0048	0.0079	0.9873	0.9866		
	2	0.0040	0.0043	0.9898	0.9934		
	1	0.0038	0.0042	0.9902	0.9939		

Figure 1: Behaviour of the regression and correlation coefficients (obtained from a GLPFM with three common latent factors) through the EM iterations

$$f_{1t} = \gamma_0 + \gamma_{11} f_{1t}^{(e)} + \gamma_{21} f_{2t}^{(e)} + \gamma_{31} f_{3t}^{(e)} + \nu_{1t}$$

$$f_{2t} = \gamma_0 + \gamma_{12} f_{1t}^{(e)} + \gamma_{22} f_{2t}^{(e)} + \gamma_{32} f_{3t}^{(e)} + \nu_{2t}$$

$$f_{3t} = \gamma_0 + \gamma_{13} f_{1t}^{(e)} + \gamma_{23} f_{2t}^{(e)} + \gamma_{33} f_{3t}^{(e)} + \nu_{3t}$$

It can be seen here that the coefficients γ_{11}, γ_{22} and γ_{33} are very close to one, whereas γ_0, γ_{21}, γ_{31}, γ_{12}, γ_{32}, γ_{13} and γ_{23} converge to zero. From Figure 1, it can be seen also that the correlations of the estimated factors with the simulated ones converge to one ($r_{f_1 f_1^{(e)}}$, $r_{f_2 f_2^{(e)}}$, $r_{f_3 f_3^{(e)}} > 90\%$ whereas $r_{f_1 f_2^{(e)}} \approx r_{f_2 f_1^{(e)}} \approx \ldots \approx 0$).

These results are confirmed by Figure 2, which displays the scatter plots of each simulated factor on the different estimated factors overlaid with a fit line. As can be clearly seen from this figure, the scatter about the line is quite small for the simulated factors and their corresponding estimated values, which implies a strong linear relationship between them. Hence, it can be argued that the proposed local EM algorithm works well and gives a good estimation of the latent factors (estimated factors very close to the simulated ones).

Note finally that in this example, the coefficient value $a = 0.95$ was used for the initialisation of the working variable z required for the implementation of the Fisher scoring step. However, the choice of any value $a \in [0,1]$ does not affect in any way the estimation results. Indeed, several simulation experiments were conducted following Saidane *et al.*[37], taking various values ranging from 0.1 to 0.9 for this parameter and the results have shown that, when convergence is reached, the algorithm gives the same solution each time.

Note also that the algorithm has converged to the optimal values after nearly 15 iterations. All these results indicate that the proposed EM–based Fisher scoring algorithm works well, for both the factor and parameters estimation.

Standard & Poor's database
Data and preliminary analysis

The database CreditPro 6.2, provided by the Standard & Poor's Risk Solutions group, summarises financial information for 9,928 bonds belonging to different rating classes within 13 industry sectors. All the rating movements between 1st January, 1982 and 31st December, 2002 as well as the different subsectors and countries are, also specified. In the following, only US bonds, belonging to the energy, transport, consumer and media sectors, will be
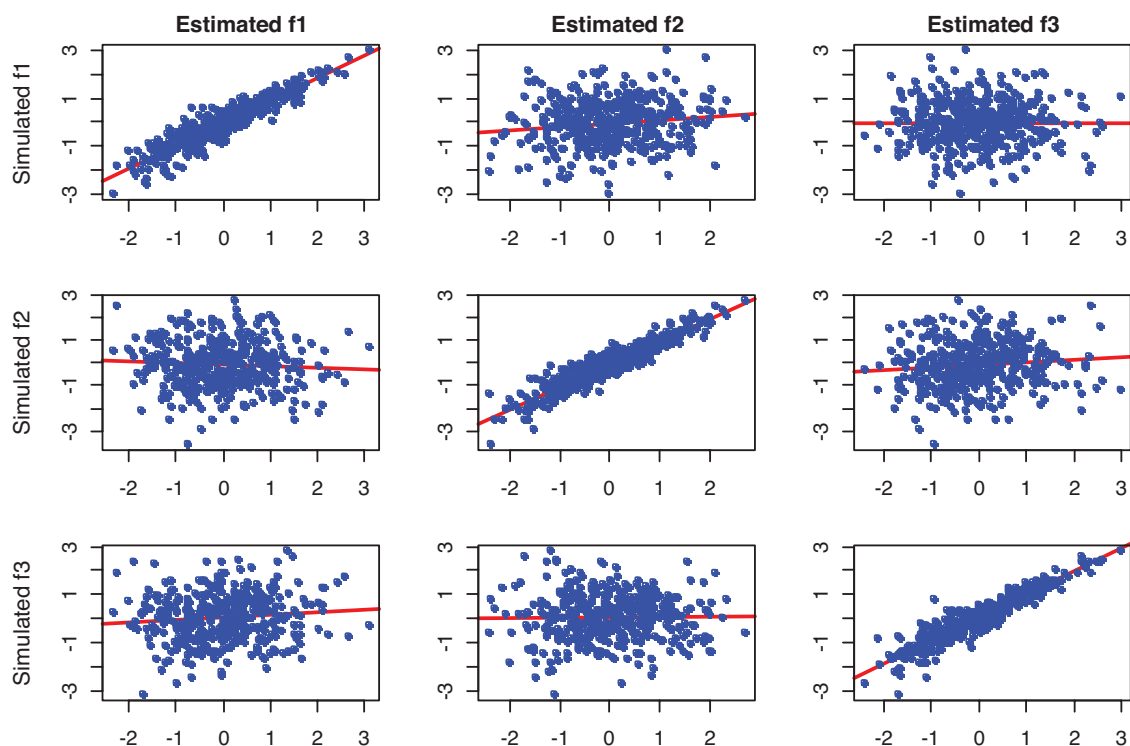
Figure 2: Scatterplots of simulated factors versus the estimated ones from a GLPFM with three common latent factors. The red line represents the simple linear regression line

considered. Figure 3 depicts the number of defaulted bonds during this period for the different considered sectors, which have been extracted from figures 2, 3, 4 and 5 in Giampieri *et al.*[38]

To assess the stability of the estimates during 1982–2002, the default events were grouped in quarterly periods. Then, the local EM algorithm was implemented to the resulting sequences, using a one-quarter rolling window framework (see Saidane[39,40]). To do this, the dataset was split into two subsets. The first one (training set) consists of 34 observations (quarters) that range from 1st January, 1982 to 30th June, 1990, on which the GLPFM was fitted. Thereafter, the defaults of the third quarter of 1990 were added and those of the first quarter of 1982 were excluded from the dataset and the model was re-estimated. This procedure was repeated to obtain means and factor loadings from 1st July, 1990 to 31st December, 2002 (50 estimates). The empirical means and standard deviations of these estimates were also computed. Note, finally, that all

these computations were made using a single-factor model in order to satisfy the identification constraints discussed in section 'Estimation in generalised linear latent factor models' by Saidane and Lavergne.[41]

Table 2 gives the total number of issued bonds during the whole period; the number of defaults for the different sectors; the empirical means of the estimated parameters and their standard deviations (into brackets) obtained from the local EM algorithm for the period 1st July, 1990 to 31st December, 2002. These results show very low standard deviations, which proves the stability of the estimates during the study period.

Backtesting results

This section aims to find the best model for estimating the credit VaR and ES using Monte Carlo simulations. To do this, the single-factor GLPFM was fitted in a first time to the observed
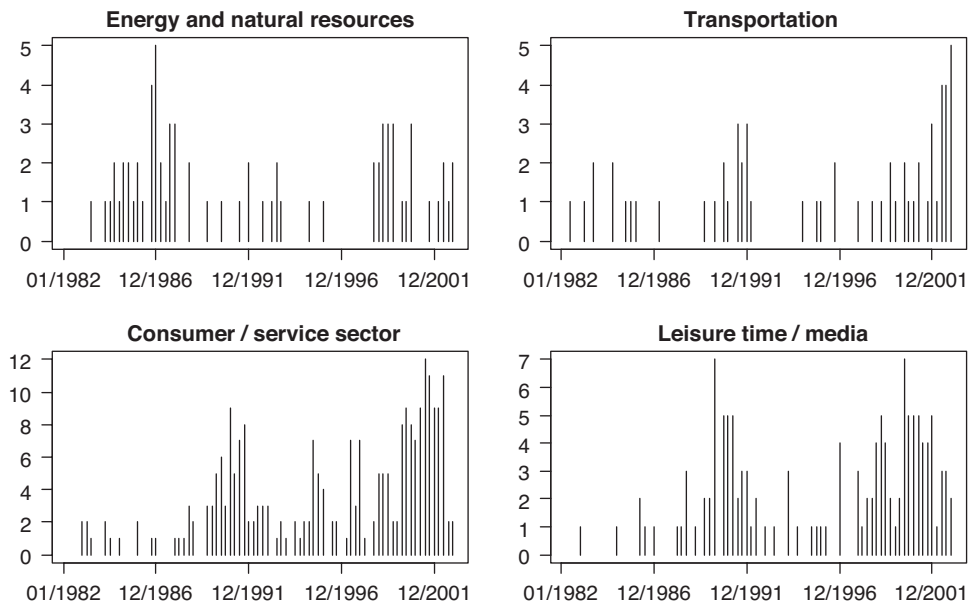
Figure 3: The number and time of defaults in each quarter for the different sectors, from January 1982 to December 2002

Table 2: Results obtained from the Standard & Poor's database

Sector	Number of bonds	Number of defaults	$\bar{\gamma}$	$\bar{\Lambda}$
Energy	420	71	-0.4423 (0.0112)	0.6762 (0.0210)
Transport	281	59	-0.2485 (0.0124)	0.6453 (0.0184)
Consumer	1041	251	1.4502 (0.0118)	0.6112 (0.0226)
Media	650	133	0.6844 (0.0122)	0.6714 (0.0231)

default sequences in order to identify the optimal latent correlation structure. Thereafter, the predictive accuracy of the model was evaluated using the one–quarter rolling window framework. The training set consists of 34 quarterly observations ranging from 1st January, 1982 to 30th June, 1990 and the backtesting period covers the range 1st June, 1990 to 31st December, 2002 (50 quarterly estimates of the VaR and ES).

To assess the precision of the credit VaR and ES estimates obtained from the model, it was assumed in the backtesting period that each bond in the loan portfolio carries one dollar. Then, a similar strategy to that developed by Saidane,[42,43] using Monte Carlo simulations from the fitted GLPFM and the CMC method was followed. After each backtesting cycle, the observed hits rates obtained by the GLPFM and the CMC approaches for the different coverage rates 2 per cent, 4 per cent, 6 per cent, 8 per cent and 10 per cent, were recorded. Moreover, the *p*-values of the unconditional coverage test developed by Kupiec[44] (UC test), the conditional coverage test and the independence test developed by Christoffersen[45] (CC and IND tests), the dynamic quantile test developed by Engle and Manganelli[46] (DQ test) and the VaR quantile test developed by Gaglianone *et al.*[47] (VQ test) were also calculated for the different coverage rates.

Table 3: Backtesting results of the GLPFM and the CMC method for the different coverage rates on the period from 30th June, 1990 to 31st December, 2002. UC, unconditional coverage test; CC, conditional coverage test; IND, independence test; DG, dynamic quantile test; VQ, VaR quantile test; AQL, average quantile loss function.

VaR	Model	% Hits	UC	IND	CC	DQ	VQ	AQL
10%	GLPFM	10%	0.5752	0.4403	0.6811	0.6327	0.5104	**0.1142**
	CMC	8%	0.0974	0.1721	0.0468	0.1743	0.0351	**0.3631**
8%	GLPFM	8%	0.4310	0.5431	0.5843	0.6109	0.4682	**0.1394**
	CMC	8%	0.0706	0.2237	0.0496	0.2076	0.0398	**0.3806**
6%	GLPFM	6%	0.4137	0.3428	0.5721	0.4759	0.4102	**0.1587**
	CMC	8%	0.0538	0.0292	0.0410	0.0506	0.0484	0.4169
4%	GLPFM	4%	0.3942	0.4842	0.6504	0.5864	0.4327	**0.1842**
	CMC	2%	0.0309	0.0356	0.0328	0.0194	0.0239	0.5243
2%	GLPFM	2%	0.3358	0.4647	0.5412	0.5462	0.3915	**0.2093**
	CMC	0%	0.0322	0.0296	0.0315	0.0136	0.0231	0.6517

Table 3 reports the hits rates: the proportions of observed defaults exceeding the VaR. The *p*-values of the backtesting VaRs are, also, reported, as well as the results given by the average quantile loss (AQL) function developed by Gonzalez-Rivera *et al.*[48] When using the adequate VaR prediction, one expects to accept the hypothesis meaning that the estimated hits rate does not differ from the expected theoretical one, and that consecutive hits are independent of each other. Note finally that methods ensuring minimum distances between the number of observed quarterly defaults and the estimated ones are always preferred.

For the coverage rates 8 per cent and 10 per cent, the GLPFM did not reject the null hypothesis indicating that the estimated hits rate is not significantly different from the expected one at 95 per cent confidence level. These results also highlight the good performance of the VaR estimates obtained by the CMC method. In this case, the *p*-values for most tests are higher than the conventional significance levels ($> 5\%$). Moreover, the results given by the model confidence set method (MCS) developed by Hansen *et al.*[49] show a set of promising models having low quantile loss values. All these results are marked in bold on Table 3. In light of these results, the single-factor GLPFM (taking into account the default correlations among the different sectors) seems to perform well and to better capture the default events, compared to the CMC method.

For the coverage rate 6 per cent, the best model is the GLPFM. The *p*-values reported by the CMC method are higher than the significance level 5 per cent only for the UC test and the DQ test. For the VaR 2 per cent and 4 per cent, the results indicate that only the GLPFM approach provided *p*-values significantly higher than 5 per cent. Hence, the proposed model seems to perform fairly well and is superior to the CMC method in the VaR prediction for the different significance levels.

Figure 4 displays the credit VaR and ES estimates obtained by the CMC and the GLPFM-based approach over the backtesting period using different significance levels. A visual inspection of Figures 4 and 5 reveals the significant influence of the latent time-varying correlations between the defaults of the different sectors, directly related to the economic situation during this period (the Savings and Loan Crisis (1980s and 1990s period) that affected the US economy for over a decade, the Gulf War recession that began in July 1990, the Balkan and South Asian Crises 1990–2002), on the behaviours of the estimated VaR and ES measures. Figure 5 shows also that the correlations obtained from the GLPFM specification are very close to the observed ones, calculated on a rolling window basis of the preceding 34 quarters. All the AQL values given in Table 3 confirm the previous results. Furthermore, these results confirm also the existence of a significant dependence relationship between the failure rate and the contagion effect in the credit
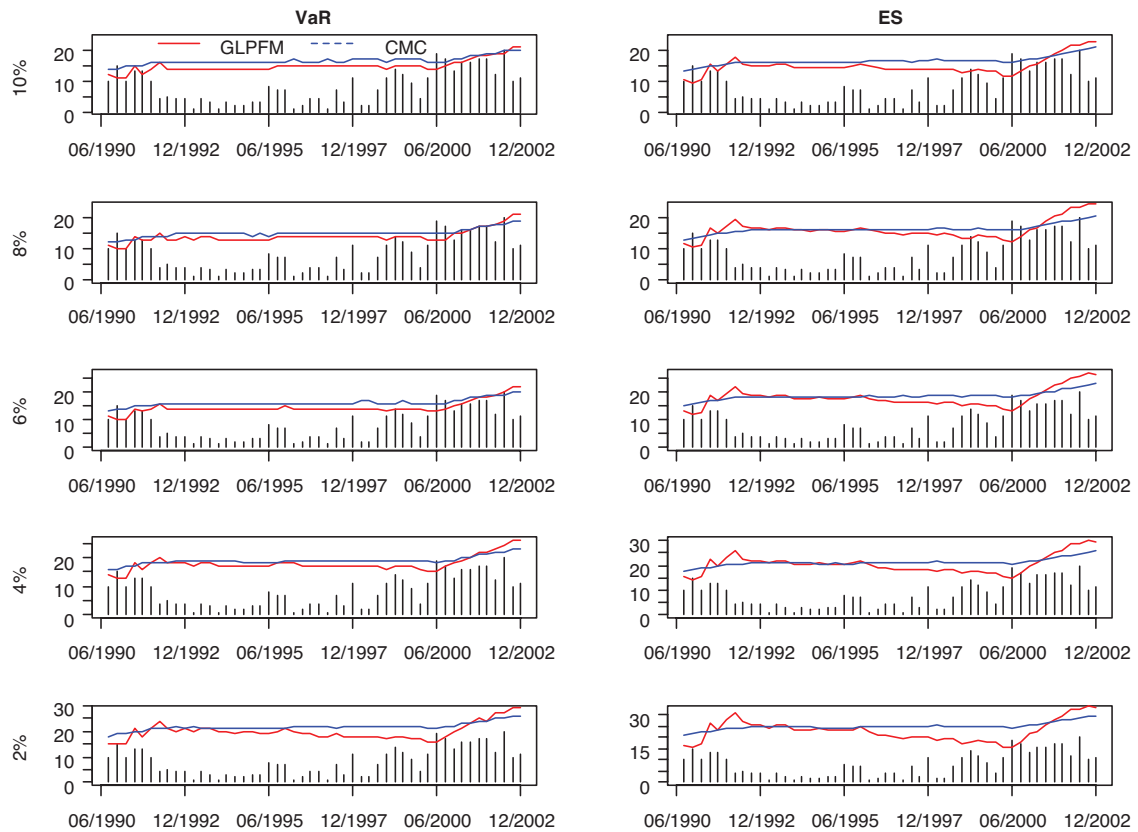
Figure 4: The credit VaR and ES values for the different significance levels over the backtesting period, from 30th June, 1990 to 31st December, 2002

market. Consequently, one can argue that the poor results obtained by the CMC approach are due to not taking into account the effects of the latent heterogeneous correlations between the defaults of the different sectors during this period. In such situations, the GLPFM may be used to quantify change in the co-movement structure of defaults over time and calibrate the simulation parameters accordingly.

Finally, Table 4 reports the ES backtesting results for the period from 30th June, 1990 to 31st December, 2002. The (McF) test developed by McNeil and Frey,[50] as well as the (NZ) and (BD) tests developed, respectively, by Nolde and Ziegel[51] and Bayer and Dimitriadis[52] were used here. In the case where the *p*-value is above the theoretical significance level (>5 per cent), then one does not reject the hypothesis that the estimated ES and the theoretical one, do not differ significantly. The McF

is used here to test the nullity of the exceedance residuals mean. The NZ and BD tests are used to test the accuracy of the estimated ES conditional to a sequence of realised defaults.

From this table, it can be observed that for the significance levels 10 per cent and 8 per cent, the corresponding *p*-values obtained from the single–factor GLPFM are significantly larger than 0.05. It can be noted also, that all the *p*-values obtained from the CMC approach are larger than 0.05, which justifies the good quality of fit to the data and the accurate ES estimation given by this method. For the coverage rates 2 per cent, 4 per cent and 6 per cent only the corresponding GLPFM's *p*-values are significantly larger than 0.05. All these results conclude that, for the different confidence levels, the GLPFM-based simulation framework gives more accurate predictions for the ES compared to those given by the CMC approach.
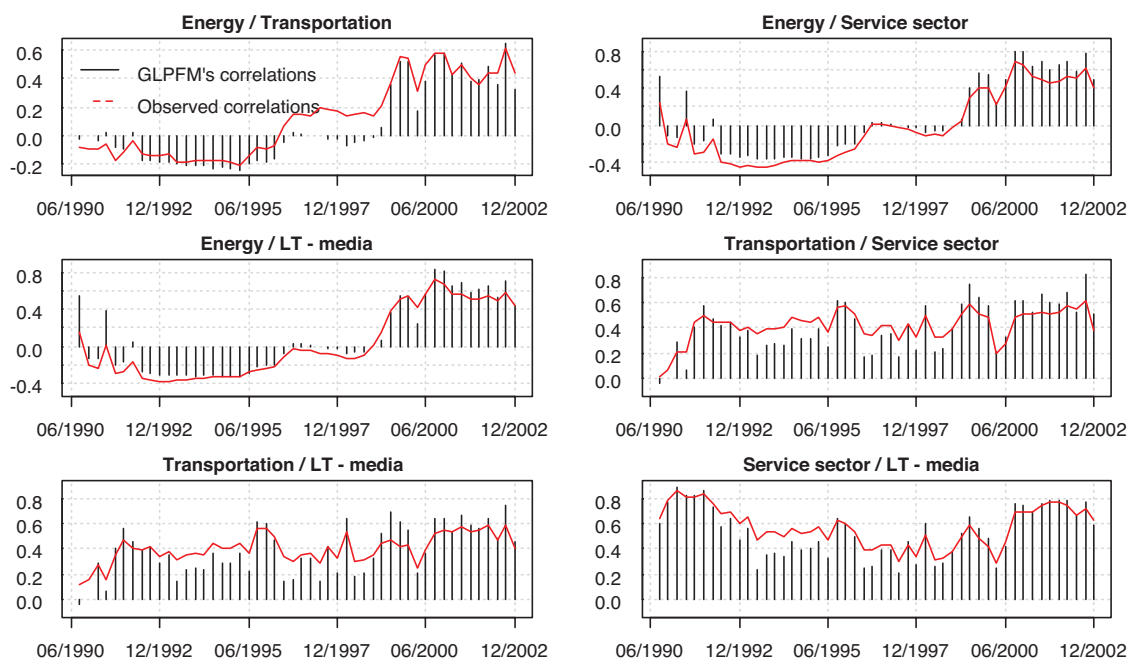
Figure 5: Default correlations between the different sectors during the backtesting period, from 30th June, 1990 to 31st December, 2002

Table 4: The ES backtesting results for the period from 30th June, 1990 to 31st December, 2002. McF, test developed by McNeil and Frey; NZ, test developed by Nolde and Ziegel; BD, test developed by Bayer and Dimitriadis.

ES	Model	McF	NZ	BD
10%	GLPFM	0.5130	0.7183	0.5072
	CMC	0.0604	0.0534	0.0512
8%	GLPFM	0.6915	0.6813	0.6764
	CMC	0.0589	0.0534	0.0521
6%	GLPFM	0.8204	0.5562	0.7906
	CMC	0.0412	0.0403	0.0319
4%	GLPFM	0.8642	0.6108	0.8511
	CMC	0.0231	0.0098	0.0244
2%	GLPFM	0.7125	0.5428	0.9237
	CMC	0.0278	0.0372	0.0151

Note finally that the accuracy of the VaR prediction results given in Table 3, can also be verified using the McF and NF tests. From these results, it can be argued that the proposed GLPFM–based simulation framework performs significantly better than the CMC approach for the different confidence levels.

CONCLUSIONS

This paper introduced a novel generalised latent Poisson factor analysis GLPFM approach for modelling credit portfolio risk with correlated defaults. Practical details of the estimation procedure, using an extended version of the EM algorithm, are also discussed. It has been shown how

to estimate the GLM parameters using the Fisher scoring algorithm, followed by a local EM step to estimate the latent predictors.

The framework is flexible enough to take into account the latent correlation structure of the default events and the market-related component in the credit risk. Thus, it can be used both in the traditional finance and risk management literature. Based on the estimation results of the GLPFM, a Monte Carlo simulation strategy had also been developed to predict the credit portfolio risk. The numerical experiments have demonstrated that the proposed approach helps to improve the credit VaR and ES predictions for different significance levels. These results may be very useful for academics concerned with the use of new intelligent statistical methods for financial engineering and risk management and for practitioners and regulators looking for modern, up-to-date techniques to improve and facilitate financial decision making.

Future studies combining this model with gaussian mixture models and hidden Markov models will be of great importance. This makes it possible to separate the market component from the total credit risk and to characterise the specific risk related to each default sequence. Further research can be conducted on (i) extending the GLPFM to a mixture of probabilistic GLPFM setting similar to the one in Mosbahi *et al.*[53] and Saidane[54]; (ii) combining the GLPFM with hidden Markov models based on the ideas in Saidane and Lavergne[55–57] and Saidane.[58,59] The local EM estimation algorithm can also be upgraded to calculate the standard errors for the GLFPM parameter estimates by the evaluation of the 'complete data variance-covariance matrix', coherently with the 'missing information principle' proposed by Louis.[60] These extensions make it possible to make comparisons between different estimation methods.

ACKNOWLEDGMENTS

References

1. Nguyen, H. and Zhou, X. (2023) 'Reduced-form Models of Correlated Default Timing: A Systematic Literature Review', *Journal of Accounting Literature*, Vol. 45, No. 1, pp. 190–205.
2. Jakob, K. (2022) 'Estimating Correlation Parameters in Credit Portfolio Models Under Time-Varying and Nonhomogeneous Default Probabilities', *Journal of Credit Risk*, Vol. 18, No. 4, pp. 29–63.
3. Cifuentes, A. and O'Connor, G. (1996) 'The Binomial Expansion Method Applied to CBO/CLO Analysis', Special Report, Moody's Investor Service, New York.
4. Frey, R. and McNeil, A. J. (2023) 'Dependent Defaults in Models of Portfolio Credit Risk', *Journal of Risk*, Vol. 6, No. 1, pp. 59–92.
5. Merton, R. C. (1974) 'On the Pricing of Corporate Debt: The Risk Structure of Interest Rates', *The Journal of Finance*, Vol. 29, No. 2, pp. 449–70.
6. Vasicek, O. (2015) 'Finance, Economics and Mathematics', John Wiley & Sons, Inc., Hoboken, New Jersey.
7. McNeil, A. J., Frey, R. and Embrechts, P. (2005) 'Quantitative Risk Management: Concepts, Techniques and Tools Princeton', Princeton University Press, New Jersey.
8. Bluhm, C., Overbeck, L. and Wagner, C. (2010) 'An Introduction to Credit Risk Modeling', 2nd edition, Chapman & Hall/CRC, New York.
9. Gupton, M., Finger, C. and Bhatia, M. (1997) 'CreditMetricsTM. Technical Document', Morgan Guaranty Trust Company Risk Management Research, New York.
10. Crosbie, P. (1999) 'Modeling Default Risk. Technical Report', KMV − LLC, San Francisco.
11. Credit Suisse Financial Products (1997) 'Credit Risk+ A Credit Risk Management Framework. Technical Report', Credit Suisse Financial Products, New York.
12. Das, S., Duffie, D., Kapadia, N. and Saita, L. (2007) 'Common Failings: How Corporate Defaults are Correlated?', *The Journal of Finance*, Vol. 62, No. 1, pp. 93–117.
13. Duffie, D., Eckner, A., Horel, G. and Saita, L. (2009) 'Frailty Correlated Default', *The Journal of Finance*, Vol. 64, No. 5, pp. 2089–2123.

14 Li, D. X. (2000) 'On Default Correlation: A Copula Function Approach', *Journal of Fixed Income*, Vol. 9, No. 4, pp. 43–54.

15 Muhajir, M. H., Six, P. and Ahn, J. H. (2022) 'Macro-Financial Determinants of Default Probability Using Copula: A Case Study of Indonesian Banks', *Bulletin of Monetary Economics and Banking*, Vol. 25, No. 4, pp. 597–622.

16 Fahrmeir, L. and Tutz, G. (2001) 'Multivariate Statistical Modelling Based on Generalized Linear Models', 2nd edition, Springer, New York.

17 McCullagh, P. and Nelder, J. A. (1989) 'Generalized Linear Models', 2nd edition, Chapman & Hall – Taylor & Francis Group, New York.

18 Saidane, M. and Lavergne, C. (2007) 'A Structured Variational Learning Approach for Switching Latent Factor Models', *Advances in Statistical Analysis*, Vol. 91, No. 3, pp. 245–68.

19 Saidane, M. and Lavergne, C. (2008) 'An EM-Based Viterbi Approximation Algorithm for Mixed–State Latent Factor Models', *Communications in Statistics – Theory and Methods*, Vol. 37, No. 17, pp. 2795–814.

20 Saidane, M. and Lavergne, C. (2009) 'Optimal Prediction with Conditionally Heteroskedastic Factor Analysed Hidden Markov Models', *Computational Economics*, Vol. 34, No. 4, pp. 323–64.

21 Saidane, M. (2017) 'A Monte-Carlo-based Latent Factor Modeling Approach with Time-Varying Volatility for Value-at-Risk Estimation: Case of the Tunisian Foreign Exchange Market', *Industrial Engineering & Management Systems*, Vol. 16, No. 3, pp. 400–14.

22 Saidane, M. (2019) 'Forecasting Portfolio-Value-at-Risk with Mixed Factorial Hidden Markov Models', *Croatian Operational Research Review*, Vol. 10, No. 2, pp. 241–55.

23 McCullagh, ref 17 above.

24 Wedel, M. and Kamakura, W. (2001) 'Factor Analysis with Mixed Observed and Latent Variables in the Exponential Family', *Psychometrika*, Vol. 66, No. 4, pp. 515–30.

25 Genton, M. G. and Ronchetti, E. (2003) 'Robust Indirect Inference', *Journal of the American Statistical Association*, Vol. 98, No. 461, pp. 1–10.

26 Moustaki, I. and Victoria-Feser, M. P. (2006) 'Bounded-influence Robust Estimation in Generalized Linear Latent Variable Models', *Journal of the American Statistical Association*, Vol. 104, No. 474, pp. 644–53.

27 Dempster, A., Laird, N. and Rubin, D. (1977) 'Maximum Likelihood from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society (Series B)*, Vol. 39, No. 1, pp. 1–38.

28 Niku, J., Brooks, W., Herliansyah, R., Hui, K. F., Taskinen, S. and Warton, D. I. (2019) 'Efficient Estimation of Generalized Linear Latent Variable Models', *PLoS ONE*, Vol. 14, No. 5, pp. 1–20.

29 Huber, P., Ronchetti, E. and Victoria-Feser, M. P. (2004) 'Estimation of Generalized Linear Latent Variable Models', *Journal of the Royal Statistical Society (Series B)*, Vol. 66, No. 4, pp. 893–908.

30 Saidane, M. (2006) 'Modèles à Facteurs Conditionnellement Hétéroscédastiques et à Structure Markovienne Cachée pour les Séries Financières', PhD Thesis, Université de Montpellier – Sciences et Techniques du Languedoc, Mathematics, Montpellier.

31 Saidane, M. (2023) 'Sequential Forecasting Strategies for Crypto Portfolio Allocation: A Dynamic Latent Factor Analysis Approach', *Journal of Administrative and Economic Sciences*, Vol. 16, No. 2, pp. 61–84.

32 Saidane, M. and Lavergne, C. (2011) 'Can the GQARCH Latent Factor Model Improve the Prediction Performance of Multivariate Financial Time Series?', *American Journal of Mathematical and Management Sciences*, Vol. 31, No. 1–2, pp. 73–116.

33 Saidane, M. (2022) 'Switching Latent Factor Value-at-risk Models for Conditionally Heteroskedastic Portfolios: A Comparative Approach', *Communications in Statistics: Case Studies, Data Analysis and Applications*, Vol. 8, No. 2, pp. 282–307.

34 Saidane, M. (2022) 'A New Viterbi-Based Decoding Strategy for Market Risk Tracking: an Application to the Tunisian Foreign Debt Portfolio during 2010–2012', *Statistika: Statistics and Economy Journal*, Vol. 102, No. 4, pp. 454–70.

35 Saidane, ref 22 above.

36 Schwarz, G. (1978) 'Estimating the Dimension of a Model', *The Annals of Statistics*, Vol. 6, No. 2, pp. 461–4.

37 Saidane, M., Bry, X. and Lavergne, C. (2013) 'Generalized Linear Factor Models: A New Local EM Estimation Algorithm', *Communications in Statistics – Theory and Methods*, Vol. 42, No. 16, pp. 2944–58.

38 Giampieri, G., Davis, M. and Rowder, M. (2005) 'Analysis of Default Data using Hidden Markov Models', *Quantitative Finance*, Vol. 5, No. 1, pp. 27–34.

39 Saidane, ref 33 above.

40 Saidane, ref 34 above.

41 Saidane, ref 32 above, pp. 78–82.

42 Saidane, ref 33 above.

43 Saidane, ref 34 above.

44 Kupiec, P. H. (1995) 'Techniques for Verifying the Accuracy of Risk Measurement Models', *The Journal of Derivatives*, Vol. 3, No. 2, pp. 73–84.

45 Christoffersen, P. F. (1998) 'Evaluating Interval Forecasts', *International Economic Review*, Vol. 29, No. 4, pp. 841–62.

46 Engle, R. F. and Manganelli, S. (2004) 'CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles', *Journal of Business & Economic Statistics*, Vol. 22, No. 4, pp. 367–81.

47 Gaglianone, W. P., Lima, L. R., Linton, O. and Smith, D. R. (2011) 'Evaluating Value-at-Risk Models via Quantile Regression', *Journal of Business & Economic Statistics*, Vol. 29, No. 1, pp. 150–60.

48 Gonzalez-Rivera, G., Lee, T. H. and Mishra, S. (2004) 'Forecasting Volatility: A Reality Check Based on Option pricing, Utility function, Value-at-Risk, and Predictive Likelihood', *International Journal of Forecasting*, Vol. 20, No. 4, pp. 629–45.

49 Hansen, P. R., Lunde, A. and Nason, J. M. (2011) 'The Model Confidence Set', *Econometrica*, Vol. 79, No. 2, pp. 453–97.

50 McNeil, A. J. and Frey, R. (2000) 'Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach', *Journal of Empirical Finance*, Vol. 7, No. (3–4), pp. 271–300.

51 Nolde, N. and Ziegel, J. F. (2017) 'Elicitability and Backtesting: Perspectives for Banking Regulation', *The Annals of Applied Statistics*, Vol. 11, No. 4, pp. 1833–74.

52 Bayer, S. and Dimitriadis, T. (2022) 'Regression-Based Expected Shortfall Backtesting', *Journal of Financial Econometrics*, Vol. 20, No. 3, pp. 437–71.

53 Mosbahi, M. N., Saidane, M. and Messabeb, S. (2017) 'Mixture of Probabilistic Factor Analyzers for Market Risk Measurement: Empirical Evidence from the Tunisian Foreign Exchange Market', *Risk Governance and Control: Financial Markets & Institutions*, Vol. 7, No. 2, pp. 158–69.

54 Saidane, ref 22 above.

55 Saidane, ref 18 above.

56 Saidane, ref 19 above.

57 Saidane, ref 20 above.

58 Saidane, ref 33 above.

59 Saidane, ref 34 above.

60 Louis, T. A. (1982) 'Finding the Observed Information Matrix when Using the EM Algorithm', *Journal of the Royal Statistical Society (Series B)*, Vol. 44, No. 2, pp. 226–33.