GRANULARITY IN QUALITATIVE FACTOR MODEL

C., GOURIEROUX (1) and A., MONFORT(2)

(January, 2009, revised September, 2009)

---

1CREST, and University of Toronto.
2CNAM and CREST.

This study benefited from financial support of the Natural Sciences and Engineering Council of Canada (NSERC), and of the Chair AXA/Risk Foundation : "Large Risks in Insurance".
Granularity in Qualitative Factor Model

ABSTRACT

This paper provides a unified setting for factor models applied to panels of qualitative observations. This setting includes as special cases the single risk factor model and its multiple factor extensions used in credit risk analysis, the stochastic migration models used for rating dynamics, and the factor models for prospective mortality tables. We consider the behavior of these models when the cross-sectional dimension is large and derive the granularity adjustments for the maximum likelihood estimators of the factor sensitivities. It is the necessary step before analyzing the effect of estimation risk on measures of credit portfolio risk. The methodology is illustrated by a Monte-Carlo study of the finite sample properties of the estimators.

Keywords: Credit Risk, Risk Concentration, Longevity, Inequality, Granularity, Factor Model.

JEL Number: 33
1 Introduction

The measures of credit portfolio risk are very sensitive to the magnitude of default correlations. The default correlations are usually determined by means of the single risk factor (SRF) model initially introduced by Vasicek (1991) [see also Tarashev, Zhu (2007), Gourieroux, Jasiak (2008)] and mandatory for credit risk analysis in the new Basel 2 regulation [see BCBS (2001)]. The model assumes that the obligors can be partitioned into homogenous classes, called cohorts further on, and that in each cohort \( k, k = 1, \ldots, K \), the individual default indicators \( Y_{i,k,t}, i = 1, \ldots, N_{kt} \) are such that:

\[
P[Y_{ikt} = 1|F_t] = \Phi \left[ \frac{\Phi^{-1}(PD_k) - kF_t}{\sqrt{1 - \rho_k^2}} \right],
\]

where \( \Phi \) is the cumulative distribution function of the standard normal distribution, \( F_t \) is a common factor, \( PD_k \) is the unconditional probability of default and \( \rho_k^2 \) is the latent correlation of default. The factor values are assumed independent and identically normally distributed:

\[
F_t \sim \text{IIN}(0,1).
\]

The aim of this paper is to provide a unified framework for this type of qualitative factor models. This framework allows for various extensions, that are, more than two alternatives for the individual qualitative variables, multi-factor models, serial dependence between factor values, cohort specific effects, logit as well as probit models. We focus on the estimation of the parameters of these qualitative factor models for large cross-sectional dimensions \( N_{kt}, k = 1, \ldots, K, t = 1, \ldots, T \) and the so-called granularity adjustment [see Gordy (2003), (2004), Gordy, Lutkebohmert (2007)] is calculated explicitly.

To get a unified setting the basic SRF model has to be rewritten in a more convenient form. More precisely, equation (1.1) is rewritten as:

\[
P_t(Y_{ikt} = 1) = \Phi(a_{kt}), \quad (1.3)
\]

where

\[
a_{kt} = \alpha_k + \beta_k F_t, \quad (1.4)
\]

\(^3\)Given the rare occurrence of joint default for typical pair of borrowers, it is practically infeasible to estimate its pairwise default correlation. To overcome the challenge, we can divide borrowers into homogenous groups of similar characteristics” [Zhang, Zhu, Lee (2008)].
with $\alpha_k \equiv \Phi^{-1}(PD_k)/\sqrt{1-\rho_k^2}$, $\beta_k = -\rho_k/\sqrt{1-\rho_k^2}$.

The specification (1.3)-(1.4) distinguishes the canonical factors $(a_{kt})$, $k = 1, \ldots, K$, valued in $\mathbb{R}$, and the fact that these canonical factors are driven by a single underlying factor $F_t$. The canonical factors as well as parameters $\alpha_k, \beta_k$ can take any real value.

We can also consider (1.2)-(1.4) as a parametric specification of the distribution of the canonical factors. This specification can be written as (1.5)-(1.7), where:

$$P_t(Y_{i,k,t} = 1) = \Phi(a_{k,t}),$$

with:

$$a_t = (a_{1,t}, \ldots, a_{K,t})' \sim \mathcal{N}[\mu(\theta), \Omega(\theta)],$$

$$\theta = (\alpha', \beta_1', \ldots, \beta_K'), \alpha = (\alpha_1, \ldots, \alpha_K)', \beta = (\beta_1, \ldots, \beta_K),'$$

$$\mu(\theta) = \alpha, \Omega(\theta) = \beta' \beta$$  

The qualitative factor model (QFM) introduced in Section 2 is an extension of the SRF model described above. The QFM is defined in two steps as follows: in the first step, the individual observations are driven by canonical time varying factors $(a_{kt})$; the second step specifies the parametric distribution of these factors. In this section and the following one the factors are assumed serially independent. Then we consider the maximum likelihood estimator of the unknown parameter $\theta$, when the cross-sectional dimensions are infinite [the so-called cross-sectionally asymptotic (CSA) estimator], and the adjustment of the estimator for large, but finite cross-sectional dimensions [i.e. the so-called granularity adjustment (GA)].

The expressions of the CSA estimator and its granularity adjustment can have an explicit form. This question is examined in Section 3, in which we consider the dichotomous qualitative models, polytomous logit models, sequential qualitative models, or log-linear probability models with general and cohort specific factors.

Different extensions are considered in Section 4, such as the introduction of dynamic factors, or the distinction between marginal and cross factors.
These extensions require numerical algorithms to find the maximum likelihood CSA estimates and their granularity adjustments. The model is implemented in Section 5, where we discuss the finite sample properties of the CSA and GA estimators. Section 6 concludes. The proofs are gathered in Appendices.

2 The qualitative factor model and its estimation

2.1 The model

The model is defined in two steps.

i) The qualitative individual observations are denoted \( Y_{i,k,t} \), where \( t = 1, \ldots, T \) denotes the date, \( k = 1, \ldots, K \) the cohort, and \( i = 1, \ldots, N_{k,t} \) the individual in the cohort. The vectors of observations \( Y_{i,k,t} \) are \( J \)-dimensional. The observations are assumed independent conditionally on the stochastic parameters \( a_{k,t} \), \( k = 1, \ldots, K, t = 1, \ldots, T \):

\[
P(Y_{i,k,t} = y_{i,k,t} | a_t) \sim f(y_{i,k,t}; a_{k,t}),
\]

where \( f \) is the elementary probability associated with the observed alternative. The stochastic parameter \( a_{k,t} \) is valued in \( \mathbb{R}^S \).

ii) The model is completed by specifying the distribution of the stochastic parameters. The stochastic parameters \( a_t = (a_{1,t}', \ldots, a_{K,t}')' \) are called canonical factors and are assumed independent and identically normally distributed:

\[
a_t \sim \mathcal{IIN}[\mu(\theta), \Omega(\theta)],
\]

where \( \theta \) is a \( p \)-dimensional unknown parameter.

Thus, the individual qualitative observations \( (Y_{i,k,t}) \) are driven by the canonical factors \( (a_{k,t}) \). In this section these factors are static, that is, they feature no serial dependence. The factor specification above is valid for panel data, where \( N_{k,t} = N_k \), independent of \( t \), and the index \( i \) represents a same individual at all dates. However, since the individual data corresponding to a given class and date are conditionally i.i.d. and since we have independence between the conditional behaviour of individuals in categories \( (k,t), (k,t^*) \),
the factor specification can also be applied to situations, where the index \( i \) in categories \((k, t)\) and \((k, t^*)\) do not correspond to the same individual. This interpretation is important in the usual SRF model used in credit risk, where default is an absorbing state and \( k \) can represent the last rating of a firm [see e.g. Gagliardini, Gourieroux (2005), Feng, Gourieroux, Jasiak (2008)].

The likelihood function of factor model (2.1)-(2.2) is:

\[
l(\theta) = \prod_{t=1}^{T} \int \ldots \int (\prod_{k=1}^{K} \prod_{i=1}^{N_{k,t}} f(y_{i,k,t}; a_{k,t}) \\
\left\{ \frac{1}{(2\pi)^{SK/2}|\text{det}\Omega(\theta)|^{1/2}} \exp\left\{ -\frac{1}{2}[a_t - \mu(\theta)]' \Omega(\theta)^{-1}[a_t - \mu(\theta)] \right\} \right\} (\prod_{k=1}^{K} da_{k,t}),
\]

where \( da_{k,t} = da_{1,k,t} \ldots da_{S,k,t} \).

This likelihood involves SK dimensional integrals. This likelihood function can be numerically optimized with respect to the unknown parameters; several algorithms exist and are implemented in statistical software packages [see e.g. Frey, McNeil (2003), Hamerle, Rosch (2005), (2007)]. However, this approach in general computationally difficult. The aim of the two following sections is to derive tractable approximations of this likelihood which are valid when i) the cross-sectional dimensions \( N_{k,t} \) are infinite, and ii) the cross-sectional dimensions are large, but finite.

### 2.2 The fixed effect maximum likelihood estimator.

The fixed effect maximum likelihood estimators provide approximations of the stochastic parameters, when distributional restrictions (2.2) are not taken into account. These estimators are defined by

\[
\hat{a}_{k,t} = \arg \max_{a_{k,t}} \sum_{i=1}^{N_{k,t}} \log f(y_{i,k,t}; a_{k,t}). \tag{2.3}
\]

The fixed effect estimator is cross-sectionally computed for each date and cohort. Under standard regularity conditions, including the cross-sectional identifiability of \( a_{k,t} \), when \( N_{k,t} \to \infty \), the estimators \( \hat{a}_{k,t} \) are independent, converge to the true values of the parameters \( a_{k,t} \) and the estimation errors are asymptotically normal:
\[ \sqrt{N_{k,t}}(\hat{a}_{k,t} - a_{k,t}) \xrightarrow{d} N[0, \Sigma_{kt}], \] 

where:

\[ \Sigma_{k,t} = \left( E \left[ -\frac{\partial^2 \log f(Y_{i,k,t}; a_{k,t})}{\partial a \partial a'} \bigg| a_{k,t} \right] \right)^{-1}. \]  

Moreover, the asymptotic variance-covariance matrix \( \Sigma_{k,t} \) of \( \hat{a}_{k,t} \) can be approximated by:

\[ \hat{\Sigma}_{k,t} = \left( -\frac{1}{N_{k,t}} \sum_{i=1}^{N_{k,t}} \frac{\partial^2 \log f(y_{i,k,t}; \hat{a}_{k,t})}{\partial a \partial a'} \right)^{-1}. \]  

These asymptotic results have to be interpreted carefully. Indeed, the fixed effect ML estimator \( \hat{a}_{k,t} \) (resp. the estimated asymptotic variance-covariance matrices) tends to a limit \( a_{kt} \) (resp. \( \Sigma_{kt} \)), which is stochastic by Assumption (2.2). It is the so-called infinitely fine grained condition [BCBS (2001), Gordy (2003)], which allows to apply the Law of Large Number and Central Limit Theorem conditionally on the canonical factors, that is, cross-sectionally for each cohort and date.

### 2.3 The cross-sectional asymptotic (CSA) estimator

Let us first consider infinite cross-sectional dimensions \( N_{k,t}, k = 1, \ldots, K, t = 1, \ldots, T \). The estimated canonical factors \( \hat{a}_{k,t} \) are sufficient summary statistics of individual observations \( (Y_{i,k,t}) \), and we can focus on the log-likelihood function corresponding to the \( (a_{k,t}) \), in which the \( (a_{k,t}) \) are replaced by the \( (\hat{a}_{k,t}) \). This log-likelihood function is given by:

\[ L_{\infty}(\theta) \propto -\frac{KST}{2} \log(2\pi) - \frac{T}{2} \log \det \Omega(\theta) - \frac{1}{2} \sum_{t=1}^{T} [\hat{a}_t - \mu(\theta)]' \Omega(\theta)^{-1} [\hat{a}_t - \mu(\theta)]. \]  

(2.7)

The CSA maximum likelihood estimator of \( \theta \) is:

\[ \hat{\theta} = \text{arg max}_{\theta} L_{\infty}(\theta). \]  

(2.8)

If the time dimension \( T \) is large \( (T \to \infty) \), we can apply the standard asymptotic theory with respect to time and deduce that \( \hat{\theta} \) converges to the true value of parameter \( \theta \), and is asymptotically normal.
\( \sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, [\text{plim} - \frac{1}{T} \frac{\partial^2 L_\infty(\hat{\theta})}{\partial \theta \partial \theta'}]^{-1}). \)  \hspace{1cm} (2.9)

### 2.4 Granularity adjustment (GA)

When the sizes \((N_{k,t})\) are large, but finite, the statistics \(\hat{a}_{k,t}\) are still sufficient summary statistics, but differ from the true stochastic parameters \((a_{k,t})\). From (2.3), we deduce that:

\[
\hat{a}_{k,t} = a_{k,t} + \frac{1}{\sqrt{N_{k,t}}} \Sigma_{k,t}^{1/2} v_{k,t} + o\left(1/N_{k,t}^{1/2}\right), \hspace{1cm} (2.10)
\]

where the errors \((v_{k,t})\) are standard normal variables, independent of each other and independent of the \(a_{k,t}\). Thus, we get approximately:

\[
\hat{a}_t \approx N[\mu(\theta), \Omega(\theta) + \hat{\Sigma}_N], \hspace{1cm} (2.11)
\]

where \(\hat{\Sigma}_N = \text{diag} [\hat{\Sigma}_{k,t}/N_{k,t}]\).

The log-likelihood function can now be approximated by:

\[
L_N(\theta) = -\frac{KST}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log \det[\Omega(\theta) + \hat{\Sigma}_N] \hspace{1cm} (2.12)
- \frac{1}{2} \sum_{t=1}^{T} [\hat{a}_t - \mu(\theta)]' [\Omega(\theta) + \hat{\Sigma}_N]^{-1} [\hat{a}_t - \mu(\theta)].
\]

The corresponding maximum likelihood estimator of \(\theta\) is defined by:

\[
\hat{\theta}_N = \arg\max_\theta L_N(\theta). \hspace{1cm} (2.13)
\]

It is proved in [Gagliardini, Gourieroux (2008)] that under standard regularity conditions, if \(N_{k,t} = \nu_{k,t} N, k = 1, \ldots, K, t = 1, \ldots, T, \) with fixed \(\nu_{k,t}\), and if \(N \to \infty, T \to \infty\) such that \(N/T \to \infty\), the estimator \(\hat{\theta}_N\) is an asymptotically efficient estimator of parameter \(\theta\).

The log-likelihood function and the ML estimator can be written as:

\[
L_N(\theta) - L_\infty(\theta) = \delta L_N(\theta) + o(1/N),
\]

\[
\hat{\theta}_N - \hat{\theta} = \delta \hat{\theta}_N + o(1/N).
\]
Their differences at order \(1/N\), i.e. \(\delta_{L_N}(\theta)\) and \(\hat{\delta}_N\), are granularity adjustments for the log-likelihood and ML estimator, respectively. The granularity adjustments are not unique, since they are defined up to a term negligible with respect to \(1/N\). Section 3 shows that it is often possible to get granularity adjustments with closed form expressions.

3 Closed form expressions

The estimates \(\hat{\theta}_N\) and \(\hat{\theta}\) can be computed numerically, but their implementation is much easier, if estimators \(\hat{a}_{k,t}\) and \(\hat{\Sigma}_{k,t}\) admit explicit expressions. This is the case for the basic qualitative factor models. We first explain how to derive an estimator \(\hat{\theta}_N\) from the CSA estimator \(\hat{\theta}\). Then, we discuss the computation of the \(\hat{a}_{k,t}'s\) and the \(\hat{\Sigma}_{k,t}'s\) for specific models (2.1) and (2.2).

3.1 Expression of \(\hat{\theta}_N\)

It is easily shown [see e.g. Gourieroux, Jasiak (2008), or Appendix 1] that:

\[
\hat{\theta}_N = \theta + \left[ -\frac{\partial^2 L_{\infty}(\hat{\theta})}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial (L_N - L_{\infty})}{\partial \theta} (\hat{\theta}) + o(1/N), \tag{3.1}
\]

or equivalently:

\[
\hat{\theta}_N = \theta + \hat{Q} \frac{\partial (L_N - L_{\infty})}{\partial \theta} (\hat{\theta}) + o(1/N), \tag{3.2}
\]

where \(\hat{Q} = \hat{V}_{as}(\hat{\theta})\) is the estimated variance-covariance matrix of the CSA maximum likelihood estimator. The term \(\frac{\partial (L_N - L_{\infty})}{\partial \theta} (\hat{\theta})\) is computed in Appendix 2. We get the following result:

**Proposition 1:** Using the notations \(\hat{\Omega} = \Omega(\hat{\theta}), \hat{\epsilon}_t = \hat{\Omega}^{-1} [\hat{a}_t - \mu(\hat{\theta})]\) and \(l_N(\theta) = L_N(\theta) - L_{\infty}(\theta)\) for the component of \(L_N(\theta) - L_{\infty}(\theta)\) associated with the observation at date \(t\), we get:

\[
\hat{\delta}_N = \hat{Q} \sum_{t=1}^{T} \frac{\partial l_N(\hat{\theta})}{\partial \theta} + o(1/N),
\]

where:
\[
\frac{\partial l_{Nt}(\hat{\theta})}{\partial \theta} = - \frac{\partial \tilde{\mu}'(\hat{\theta})}{\partial \theta} \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\epsilon}_t + \sum_{j=1}^{KS} \sum_{l=1}^{KS} \left[ \frac{1}{2} \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\epsilon}_t - \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\epsilon}_t' \right]_{jl} \frac{\partial \Omega_{jl}(\hat{\theta})}{\partial \theta}.
\]

Moreover the asymptotic variance of the estimator \( \hat{\theta}_N \) is consistently estimated by:

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial L_{Nt}(\hat{\theta})}{\partial \theta} \frac{\partial L_{Nt}(\hat{\theta})}{\partial \theta'} = \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial L_{\infty t}(\hat{\theta})}{\partial \theta} + \frac{\partial l_{Nt}(\hat{\theta})}{\partial \theta} \right] \left[ \frac{\partial L_{\infty t}(\hat{\theta})}{\partial \theta} + \frac{\partial l_{Nt}(\hat{\theta})}{\partial \theta} \right],
\]

where \( \frac{\partial l_{Nt}(\hat{\theta})}{\partial \theta} \) is given above and

\[
\frac{\partial L_{\infty t}(\hat{\theta})}{\partial \theta} = \frac{\partial \tilde{\mu}'(\hat{\theta})}{\partial \theta} \hat{\epsilon}_t + \frac{1}{2} \sum_{j=1}^{KS} \sum_{l=1}^{KS} \left[ \hat{\epsilon}_{jt} \hat{\epsilon}_{lt} - (\hat{\Omega}^{-1})_{jl} \right] \frac{\partial \Omega_{jl}(\hat{\theta})}{\partial \theta}.
\]

### 3.2 Simple cross-sectional models

The qualitative cross-sectional models (2.1) used in practice have often simple forms, and provide explicit expressions for the estimated canonical factors \( \hat{a}_{kt} \) and their asymptotic variance-covariance matrices \( \hat{\Sigma}_{k,t}'s. \) Loosely speaking, the standard parametric qualitative models, such as logit or probit, have just to be transformed to ensure that their parameters are valued in \( IR^S. \) Several examples are given below.

i) The dichotomous qualitative model

The qualitative variable takes value 0, 1 and is such that:

\[
Y_{i,k,t} \sim B[1, G(a_{k,t})],
\]

(3.3)
where $G$ is a cumulative distribution function (cdf), which is the Gaussian cdf for the probit model or the logistic cdf for the logit model. The dichotomous probit model is used in the risk factor model recommended by the Basel Committee [BCBS (2001)], in the factor models proposed by the specialized agencies such as KMV [see the Global Correlation Model described in Crosbie, Bohn (2003), Crosbie (2005)], or CreditMetrics [see Finger (1999) and the restricted version in Gordy (2000)], and more generally in all models derived from the SRF model initially introduced by Vasicek (1991) [see e.g. Frey, McNeil, Nyfeler (2001)]. The dichotomous logit model is used in McKinsey’s Credit Portfolio View [Wilson (1997) a,b].

In this framework, we get:

$$
\hat{a}_{k,t} = G^{-1}(\hat{PD}_{k,t}),
$$

where $\hat{PD}_{k,t}$ is the observed frequency of alternative 1 per cohort and date, that is, the default frequency if alternative 1 represents default, and

$$
\hat{\Sigma}_{k,t} = \frac{G(\hat{a}_{k,t})[1 - G(\hat{a}_{k,t})]}{[g(\hat{a}_{k,t})]^2},
$$

where $g$ denotes the density of distribution $G$.

ii) The multinomial logit model

The qualitative variable can take $S + 1$ alternatives $s = 0, 1, \ldots, S$, and the elementary probabilities are:

$$
\begin{align*}
p_{0,k,t} &= P[Y_{i,k,t} = 0] = (1 + \exp a_{1,k,t} + \ldots + \exp a_{S,k,t})^{-1}, \\
p_{s,k,t} &= P[Y_{i,k,t} = s] = \exp(a_{s,k,t})(1 + \exp a_{1,k,t} + \ldots + \exp a_{S,k,t})^{-1}
\end{align*}
$$

for $s \geq 1$

The multinomial logit model (MLN) is typically used for describing qualitative choice behavior by individuals [see e.g. McFadden (1973), (1976)]. It is appropriate for modelling the demand of financial assets by households, that is the way they allocate their wealth between the main classes of assets $s = 0, 1, \ldots, S$, such as saving account, pension fund, life insurance, bonds, or stocks. The demand uncertainty is a major component when studying the
risk of the portfolio of life insurance contracts holds by an insurance company in the context of Solvency 2.

Let us denote by \( \hat{p}_{s,k,t} \) the observed frequency of alternative \( s \) for a given cohort \( k \) and date \( t \). The fixed effect ML estimator of \( a_{s,k,t}, s = 1, \ldots, S \) is the log-odd ratio:

\[
\hat{a}_{s,k,t} = \log(\hat{p}_{s,k,t}/\hat{p}_{0,s,t}), s = 1, \ldots, S, k = 1, \ldots, K, t = 1, \ldots, T. \tag{3.8}
\]

The associated variance-covariance matrix of the estimated transformed parameter is:

\[
\Sigma_{k,t} = G_{k,t} Q_{k,t} G_{k,t}',
\]

where \( Q_{k,t} = \text{diag}(p_{k,t}) - p_{k,t} p_{k,t}', \) where \( p_{k,t}' = (p_{0,k,t}, p_{1,k,t}, \ldots, p_{S,k,t}) \),

\[
G_{k,t} = \begin{pmatrix}
-1/p_{0,k,t} & 1/p_{1,k,t} & 0 & \cdots & 0 \\
\vdots & 0 & 1/p_{2,k,t} & \ddots & \vdots \\
-1/p_{0,k,t} & 0 & \cdots & \cdots & 1/p_{S,k,t}
\end{pmatrix}.
\]

We get:

\[
\Sigma_{k,t} = (1/p_{0,k,t}) e_S e_S' + \text{diag} (1/\hat{p}_{k,t}),
\]

where \( e_S = (1, \ldots, 1)' \) and \( \hat{p}_{k,t} = (p_{1,k,t}, \ldots, p_{S,k,t})' \).

**iii) The sequential logit model**

Let us consider a sequential logit model with three alternatives and omit index \( k \) for expository purpose. The elementary probabilities can be written as:

\[
p_1 = \frac{1}{1 + \exp(a_1)}, \quad p_2 = \frac{\exp(a_1)}{1 + \exp(a_1)} \frac{1}{1 + \exp(a_2)}, \\
p_3 = \frac{\exp(a_1)}{1 + \exp(a_1)} \frac{\exp(a_2)}{1 + \exp(a_1)} \frac{1}{1 + \exp(a_2)}.
\]

Such a model can be used in a preliminary joint analysis of default and recovery. Indeed, a loan can be declared in default for different reasons,
for instance if the borrower has not satisfied his/her contractual monthly payment, but also if the credit institution expects future difficulties, even if the borrower continues to satisfy his/her contractual obligations. These different reasons for default can be distinguished by means of the recovery rate, which is close to 1 in the second case. The three alternatives in this application could be:

- alternative 1: no default,
- alternative 2: default with high recovery rate, larger than 95%, say.
- alternative 3: default with smaller recovery rate.

The dichotomous logit model with elementary probability \( p/(1+\exp(a_1)) \) is introduced to analyze default occurrence, whereas the dischomous logit model with elementary probability \( 1/[1 + \exp(a_2)] \) will capture the high recovery level.

For the sequential logit model, we get:

\[
a_1 = \log \frac{1-p_1}{p_1}, a_2 = \log(p_3/p_2),
\]

that correspond to marginal and conditional log-odd ratios, respectively.

The asymptotic distribution of the estimated transformed parameters is:

\[
\Sigma = GQG',
\]

where \( Q = diagp = pp' \),

\[
G = \begin{pmatrix}
- \frac{1}{p_1(1-p_1)} & 0 & 0 \\
0 & -1/p_2 & 1/p_3 \\
1/p_2 & 1/p_3 \\
\end{pmatrix}.
\]

We get: \( \Sigma = \begin{pmatrix}
\frac{1}{p_1(1-p_1)} & 0 & 0 \\
0 & 1/p_2 & 1/p_3 \\
1/p_2 & 1/p_3 \\
\end{pmatrix}.\)

iv) The log-linear probability model for contingency tables

Alternative parametrizations have also been proposed for multidimensional contingency tables with focus on dependence measures [see e.g. Koenig, Nerlove, Oudiz (1979), Maddala, Trost (1981), Nerlove (1983), Nerlove, Press (1986)]. The log-linear probability models propose decompositions of the log-probabilities to distinguish the overall, specific and cross effects of the
alternatives. For illustration, let us consider a $2 \times 2$ contingency table with elementary probabilities: $p_{1,1,t}, p_{1,0,t}, p_{0,1,t}, p_{0,0,t}$, say.

This type of model can be used to follow the investment versus speculative ratings of the short and long term corporate debts. The alternatives are:

- alternative 1,1: investment rating for both short and long term debts,
- alternative 1,0: investment rating for the short term debt, speculative rating for the long term debt,
- and so on.

The new parametrization is such that:

\[
\begin{align*}
\log p_{1,1,t} &= \nu_t + a_{1,t} + a_{2,t} + a_{3,t}, \\
\log p_{1,0,t} &= \nu_t + a_{1,t} - a_{2,t} - a_{3,t}, \\
\log p_{0,1,t} &= \nu_t - a_{1,t} + a_{2,t} - a_{3,t}, \\
\log p_{0,0,t} &= \nu_t - a_{1,t} - a_{2,t} + a_{3,t},
\end{align*}
\]

or equivalently:

\[
\begin{align*}
a_{1,t} &= \frac{1}{4} \log\left(\frac{p_{1,1,t}p_{1,0,t}}{p_{0,1,t}p_{0,0,t}}\right), \\
a_{2,t} &= \frac{1}{4} \log\left(\frac{p_{1,1,t}p_{0,1,t}}{p_{0,0,t}p_{1,0,t}}\right), \\
a_{3,t} &= \frac{1}{4} \log\left(\frac{p_{1,1,t}p_{0,0,t}}{p_{0,1,t}p_{1,0,t}}\right).
\end{align*}
\]

In particular, parameter $a_{3,t}$ measures the dependence between the two qualitative variables, whose distribution is described in the contingency table, that is, the dependence between the ratings of the short and long term debts in the example above.

A simple computation shows that the asymptotic variance-covariance matrix of the estimated transformed parameters has the elements:
\[
\begin{align*}
\sigma_{1,t}^2 &= \sigma_{2,t}^2 = \sigma_{3,t}^2 = \frac{1}{16} \left( \frac{1}{p_{0,0,t}} + \frac{1}{p_{0,1,t}} + \frac{1}{p_{1,0,t}} + \frac{1}{p_{1,1,t}} \right), \\
\sigma_{1,2,t} &= \frac{1}{16} \left( \frac{1}{p_{0,0,t}} - \frac{1}{p_{0,1,t}} - \frac{1}{p_{1,0,t}} + \frac{1}{p_{1,1,t}} \right), \\
\sigma_{1,3,t} &= \frac{1}{16} \left( -\frac{1}{p_{0,0,t}} + \frac{1}{p_{0,1,t}} - \frac{1}{p_{1,0,t}} + \frac{1}{p_{1,1,t}} \right), \\
\sigma_{2,3,t} &= \frac{1}{16} \left( -\frac{1}{p_{0,0,t}} - \frac{1}{p_{0,1,t}} + \frac{1}{p_{1,0,t}} + \frac{1}{p_{1,1,t}} \right).
\end{align*}
\]

Similar approaches can be applied to the other types of log-linear probability models.

### 3.3 The model for canonical factors

**i) The restricted factor model**

The number of canonical factors, components of \(a_t\), is equal to \(K_S\) and we assume that the factor space can be restricted by considering a smaller number \(L\) of common factors and idiosyncratic error terms:

\[
a_t = \alpha + \sum_{l=1}^{L} \beta_l F_{lt} + \eta u_t, \tag{3.9}
\]

\[
= \alpha + \beta F_t + \eta u_t, \tag{3.10}
\]

where \(\alpha\) (resp. \(\beta\)) is a vector of dimension \(K_S\) [resp. a matrix of size \((K_S, L)\)]; \(\eta\) is a positive scalar, and the variables \(F_{lt}, l = 1, \ldots, L, u_t\) are independent, Gaussian variables \(F_{lt} \sim N(0,1), u_t \sim N(0, I_d)\). We deduce that:

\[
a_t \sim II N(\alpha, \beta \beta' + \eta^2 I_d). \tag{3.11}
\]

We have: \(\theta = (\alpha', \beta_1', \ldots, \beta_L', \eta^2), \mu(\theta) = \alpha, \Omega(\theta) = \beta \beta' + \eta^2 I_d. \tag{3.12}\)
As usual factor $F$ is defined up to an orthogonal linear transformation and for parameter identification the beta vectors can be assumed orthogonal:

$$\beta'_k \beta_l = 0, \forall k \neq l.$$  \hspace{1cm} (3.13)

The restricted factor model (3.11)-(3.13) extends the standard specification used by practitioners, which does not include the idiosyncratic effects, that is, assumes that $\eta^2$ is equal to zero. The restriction $\eta^2 = 0$ induces an underestimation of the risk, a mispricing of the associated derivatives and spurious arbitrage opportunity.

**ii) Derivatives of $\mu$ and $\Omega$**

The granularity adjustments involve the derivatives of the mean $\mu$ and of volatility-covolatility $\Omega$ matrix of the canonical factors. They admit explicit forms.

$$\frac{\partial \mu}{\partial \alpha} = 1d, \frac{\partial \mu}{\partial \beta_i} = 0, \frac{\partial \mu}{\partial \eta^2} = 0, \frac{\partial \Omega_{ij}}{\partial \alpha} = 0, \frac{\partial \Omega}{\partial \eta^2} = 1d,$$

and

$$\frac{\partial \Omega_{ij}}{\partial \beta_{kl}} = 0, \text{ if } k \neq i,j, = \beta_{i} \text{ if } k = j \neq i, = \beta_{i} \text{ if } k = i \neq j, = 2\beta_{i}, \text{ if } k = i = j,$$

where $\beta_{kl}$ denotes the $(k,l)$ entry of $\beta$.

**iii) CSA maximum likelihood estimator of $\theta$**

Under the restricted factor model introduced in Section 3.3 i), the CSA maximum likelihood estimator of $\theta$ admits an explicit expression (see Anderson (2003), and Appendices 3 and 4).

**Proposition 2** : Let us denote $\bar{a}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{a}_t$ and $\hat{V}_T = \frac{1}{T} \sum_{t=1}^{T} (\hat{a}_t - \bar{a}_T)(\hat{a}_t - \bar{a}_T)'$ the sample mean and sample variance-covariance matrix of the estimated canonical factors. Let us also consider the eigenvalues of $\hat{V}_T$ in decreasing order $\lambda_1 > \lambda_2 \ldots >$ and the associated eigenvectors $e_l, l = 1, \ldots, L$ with unit norm. The CSA maximum likelihood estimators of the components of parameter $\theta$ are:
\[\hat{\alpha} = \bar{a}_T,\]
\[\hat{\eta}^2 = [Tr(\hat{V}_T) - \sum_{l=1}^{L} \hat{\lambda}_l]/(KS - L),\]
\[\hat{\beta}_l = (\hat{\lambda}_l - \hat{\eta}^2)^{1/2} e_l.\]

The explicit formulas for the asymptotic variance-covariance matrix of the CSA estimators are easily derived directly, after an appropriate change of parameter \(^4\). Let us denote \(\beta^*_l = \beta_l/\eta, l = 1, \ldots, L, B^* = (\beta^*_1, \ldots, \beta^*_L),\)
\[J = \lim_{T \to \infty} -\frac{1}{T} \frac{\partial^2 L_\infty}{\partial \theta \partial \theta^t}\]
the asymptotic Hessian of the log-likelihood function and \(G\) the Jacobian matrix of the orthogonality restrictions. By applying general results on constrained maximum likelihood estimation [see Gourieroux, Monfort (1995), Chapter 10], we have for large \(T\):
\[
\sqrt{T}\begin{bmatrix}
\hat{\alpha} \\
\hat{\eta}^2 \\
\hat{\beta}^*_1 \\
\vdots \\
\hat{\beta}^*_L
\end{bmatrix}
- \begin{bmatrix}
\alpha \\
\eta^2 \\
\beta^*_1 \\
\vdots \\
\beta^*_L
\end{bmatrix}
\xrightarrow{d} N[0, J^{-1} - J^{-1}G(G'J^{-1}G)\^{-1}G'J^{-1}].
\]
We prove in Appendix 4 iv) that:
\[
J = \begin{pmatrix}
J_{\alpha \alpha} & 0 & 0 & \ldots & 0 \\
0 & J_{\eta^2 \eta^2} & J_{\eta^2 \beta_1} & \ldots & J_{\eta^2 \beta_L} \\
\vdots & J_{\beta_1 \eta^2} & J_{\beta_1 \beta_1} & 0 & 0 \\
\ldots & 0 & \ddots & \ldots & \ldots \\
0 & J_{\beta_L \eta^2} & 0 & \ldots & J_{\beta_L \beta_L}
\end{pmatrix},
\]
\(^4\)The initial derivation in Lawley, Maxwell (1971) provides only an approximated variance-covariance matrix [see also Jennrich, Thayer (1977)].
\[ J_{\alpha\alpha} = [\eta^2(Id + B^BB^\prime)]^{-1}, \]
\[ J_{\eta^2} = KS/(2\eta^4), \]
\[ J_{\eta^2\beta^*} = \beta^* /[\eta^2(1 + \beta^\prime\beta^*)], \]
\[ J_{\beta^*,\beta^*} = \frac{\beta^\prime\beta^*}{1 + \beta^\prime\beta^*}Id - \frac{B^BB^\prime}{1 + \beta^\prime\beta^*} + \frac{2\beta^\prime\beta^*}{(1 + \beta^\prime\beta^*)^2}. \]

In a single risk factor model, there is no orthogonality restriction, and the variance-covariance matrix reduces to \( J^{-1} \) [see Gourieroux-Jasiak (2008)].

In the 2-factor case, there is a single orthogonality restriction \( \beta^\prime_1\beta^*_2 = 0 \), and the Jacobian matrix is:

\[
G' = (0, 0, \beta^*_2, \beta^*_1). \]

In the 3-factor case, we get 3-orthogonality restrictions: \( \beta^\prime_1\beta^*_2 = 0, \beta^\prime_1\beta^*_3 = 0, \beta^\prime_2\beta^*_3 = 0 \). The Jacobian matrix is:

\[
G' = \begin{pmatrix}
0 & 0 & 0 & \beta^*_2 & \beta^*_3 \\
0 & 0 & \beta^*_1 & 0 & \beta^*_3 \\
0 & 0 & 0 & \beta^*_1 & \beta^*_2 \\
\end{pmatrix}. \]

4 Extensions

The qualitative factor models of Section 3 can be extended in different ways. We first consider the introduction of dynamic factors, then the possibility to distinguish marginal and cross factors. In these extensions the CSA estimators and the granularity adjustment no longer admit explicit expressions and have to be computed numerically.

4.1 Dynamic factor

A typical dynamic factor model is given by:

\[ Y_{i,k,t} \sim f(y_{i,k,t}; a_{k,t}), \quad (4.1) \]
where

\begin{align*}
a_t &= \alpha + \beta' F_t + \eta u_t, \quad (4.2) \\
F_t &= \Phi F_{t-1} + (I_d - \Phi \Phi')^{1/2} w_t, \quad (4.3) \\
    &\quad w_t \sim IIN(0, I_d), u_t \sim IIN(0, I_d).
\end{align*}

The variance-covariance innovation matrix in (4.3) ensures that the underlying factors admit an unconditional variance-covariance matrix equal to the identity. When \( \Phi = 0 \), the model reduces to the static specification considered in (3.9) - (3.11), and the unconditional and conditional analysis of risk coincide. By introducing serial dependence for the factors, we allow for a dynamic analysis of risk and risk dependence, and for computation of conditional Value-at-Risk. For instance a specific dynamic probit factor has been proposed for the analysis of sector specific default rates in Cipollini, Missaglia (2007). In this dynamic setting, the CSA model becomes:

\[
\begin{cases}
\hat{a}_t \simeq \alpha + \beta' F_t + \eta u_t, \\
F_t = \Phi F_{t-1} + (I_d - \Phi \Phi')^{1/2} w_t,
\end{cases}
\quad (4.4)
\]

whereas the granularity adjusted model becomes:

\[
\begin{cases}
\hat{a}_{k,t} \simeq a_k + \beta_k' F_t + \eta u_{k,t} + (\hat{\Sigma}_{k,t}/N_{k,t})^{1/2} v_{k,t}, k = 1, \ldots K, \\
F_t = \Phi F_{t-1} + (I_d - \Phi \Phi')^{1/2} w_t.
\end{cases}
\quad (4.5)
\]

Both CSA and GA models have a Gaussian linear state space representation, with (4.3) as the transition equation and the first equation of either (4.4), or (4.5) as the measurement equation. Therefore, even if the CSA and GA maximum likelihood estimators have no closed form expressions, they are easily computed numerically by applying the Kalman Filter included in the standard statistical software packages. Moreover, the Kalman Filter provides also the CSA and GA filtered values of factors \( F_t, t = 1, \ldots, T \), and the predictions and prediction intervals for future values \( F_{T+h}, a_{T+h} \), say. The filtered factor values are useful for factor interpretation, for instance for interpreting some of the factors in terms of business cycle.
4.2 Marginal and cross factors

Restricted factor models similar to model (3.9)-(3.10) can also be introduced with factors admitting ex-ante interpretations. More precisely, let us assume that the cohorts are doubly indexed by $k, l$, with canonical factors $a_{k,l,t}$, say. Typically, in credit risk analysis, the corporates can be segmented by country and industrial sector. In such a framework, it is interesting to distinguish between pure geographical factors, pure industrial factors and factors with cross effects. The basic factor model (3.9)-(3.10) is extended to:

$$a_{k,l,t} = \alpha_{k,l} + \beta^1_k F^1_t + \beta^2_l F^2_t + \beta^{1,2}_{kl} F^{1,2}_t + \eta u_{k,l,t},$$  \hspace{1cm} (4.6)

with $\sum_{k=1}^K \beta^1_k = 0$, $\sum_{l=1}^L \beta^2_l = 0$, $\sum_{k=1}^K \beta^{1,2}_{k,l} = 0$, $\forall t$, $\sum_{l=1}^L \beta^{1,2}_{k,l} = 0$, $\forall k$.

The components of $F^1_t, F^2_t, F^{1,2}_t, u_t$ are assumed independent standard normal variables.

Factor $F^1_t$ (resp. $F^2_t, F^{1,2}_t$) captures the marginal effects corresponding to characteristic $k$ [resp. the marginal effects corresponding to $l$, the cross effects].

In general the CSA estimators of parameters $\alpha, \beta^1, \beta^2, \beta^{12}$ have no closed form expressions and have to be computed numerically. However, important simplifications can arise in special cases. For instance, in a factor model

$$a_{k,l,t} = \alpha_{k,l} + \beta^1_k F^1_t + \beta^{1,2}_{kl} F^{1,2}_t + \eta u_{k,l,t},$$  \hspace{1cm} (4.7)

with a single type of marginal effect, the variance-covariance matrix $\hat{V}_T$ (see Proposition 1) can be decomposed into

$$\hat{V}_T = \hat{B} \hat{V}_T + \hat{W} \hat{V}_T,$$  \hspace{1cm} (4.8)

where $\hat{B} \hat{V}_T$ [resp. $\hat{W} \hat{V}_T$] is the variance-covariance matrix of $\hat{a}_t$ between cohort [resp within cohorts]. Then, the CSA maximum likelihood estimators of the parameters of interest can be deduced from separate spectral decompositions of $\hat{B} \hat{V}_T, \hat{W} \hat{V}_T$ [which differ from the spectral decomposition of their sum as considered in (3.3)].
5 A Monte-Carlo study

Let us now illustrate the finite sample properties of the CSA and GA estimators for a stratified dichotomous logit model with factor. The specification is chosen to ensure closed form expressions of the estimators [see Section 3.1].

5.1 The dichotomous logit model with factor

Let us consider a dichotomous logit model defined by:

\[ Y_{i,k,t} \sim B[1, G(a_{k,t})], i = 1, \ldots, N, k = 1, 2, t = 1, \ldots, T, \]  

(5.1)

where \( G \) denotes the logit transformation:

\[ G(z) = \frac{1}{1 + \exp(-z)}. \]  

(5.2)

In this model, we get [see Section 3.2 i)]:

\[ \hat{a}_{k,t} = \log[\bar{Y}_{k,t}/\{1 - \bar{Y}_{k,t}\}], \]  

(5.3)

where \( \bar{Y}_{k,t} = \frac{1}{N} \sum_{i=1}^{N} Y_{i,k,t} \), and \( \hat{\Sigma}_{k,t} = 1/\{\bar{Y}_{k,t}(1 - \bar{Y}_{k,t})\} \), \( k = 1, 2 \).

The dichotomous logit model above is completed by the factor restriction [see Section 3.3 i)]

\[ a_t = (a_{1,t}, a_{2,t})' = \alpha + \beta F_t + \eta u_t, \]  

(5.4)

where \( F_t \) is the one-dimensional common factor and \( u_t \) the 2-dimensional idiosyncratic error. The variables \( F_t, u_{1t}, u_{2t} \), are independent standard normal. Parameters \( \alpha, \beta \) are unknown two dimensional vectors, whereas the scalar \( \eta \) is assumed to be given.

5.2 The CSA estimators

By applying Proposition 2, the CSA estimator of parameter \( \alpha \) is:

\[ \hat{\alpha} = \bar{a}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{a}_t, \]  

(5.5)

where \( \hat{a}_t = (\hat{a}_{1,t}, \hat{a}_{2,t})' \), and the CSA estimator of parameter \( \beta \) is:
\[
\hat{\beta} = (\hat{\lambda}_1 - \eta^2)^{1/2} e_1, \tag{5.6}
\]
where \( \hat{\lambda}_1 \) is the largest eigenvalue of the volatility matrix:

\[
\hat{V} = \frac{1}{T} \sum_{t=1}^{T} (\hat{\alpha}_T - \bar{\alpha}_T)(\hat{\alpha}_t - \hat{\alpha})', \tag{5.7}
\]
and \( e_1 \) is its associated eigenvector.

### 5.3 The GA estimators

The granularity adjusted estimator of \( \alpha \) is derived from Proposition 1. We get:

\[
\hat{\alpha}_N = \hat{\alpha} - \frac{1}{T} \sum_{t=1}^{T} (\hat{\Sigma}_N \hat{\epsilon}_t), \tag{5.8}
\]
where \( \hat{\Sigma}_N = \frac{1}{N} \begin{pmatrix} \hat{\Sigma}_{1,t} & 0 \\ 0 & \hat{\Sigma}_{2,t} \end{pmatrix}, \hat{\epsilon}_t = \hat{\Omega}^{-1}(\hat{\alpha}_t - \hat{\alpha}), \hat{\Omega} = \eta^2 I + \hat{\beta} \hat{\beta}'. \]

The granularity adjusted estimator of \( \beta \) is:

\[
\hat{\beta}_N = \hat{\beta} + \eta^2 \left[ \frac{\hat{\beta}' \hat{\beta}}{\eta^2 + \hat{\beta}' \hat{\beta}} I + \frac{\hat{\beta}' \hat{\beta}'}{\eta^2 + \hat{\beta}' \hat{\beta}} + \frac{2\eta^2 \hat{\beta}' \hat{\beta}'}{\eta^2 + \hat{\beta}' \hat{\beta}} \right]^{-1} C, \tag{5.9}
\]
where \( C = B_{11} \left( \begin{array}{cc} 2\hat{\beta}_1 \\ 0 \end{array} \right) + (B_{12} + B_{21}) \left( \begin{array}{c} \hat{\beta}_2 \\ \hat{\beta}_1 \end{array} \right) + B_{22} \left( \begin{array}{c} 0 \\ 2\hat{\beta}_2 \end{array} \right), \)

\[
B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \frac{1}{2} \hat{\Omega}^{-1} \hat{\Sigma}_N \hat{\Omega}^{-1} - \hat{\Omega}^{-1} \bar{S}_N,
\]

\[
\hat{\Sigma}_N = \frac{1}{T} \sum_{t=1}^{T} (\Sigma_N), \bar{S}_N = \frac{1}{T} \sum_{t=1}^{T} (\Sigma_N \hat{\epsilon}_t \hat{\epsilon}_t').
\]

### 5.4 The Monte-Carlo experiment

We performed a Monte Carlo experiment with the following numerical values: \( \eta^2 = .1, \alpha_1 = .3, \alpha_2 = .5, \beta_1 = .1, \beta_2 = .2, T = 10 \) and \( 50 \), and consider
different values of $N$: 50, 100, 200, 500, respectively. For each setting, we computed the CSA and GA estimates $\hat{\alpha}, \hat{\alpha}_N, \hat{\beta}, \hat{\beta}_N$, and we replicated 5000 times the whole simulation experiment. The empirical biases, standard errors, root mean square errors of the different estimators are given in Tables 1 and 2 and their empirical probability density functions in Figures 1-8.

Tables 1 and 2 and Figures 1-8 show that the efficiency gain obtained from the granularity adjustment decrease with $N$, and that this gain is moderate for parameter $\alpha_1, \alpha_2$, but very important for parameters $\beta_1, \beta_2$. Note that these latter parameters are crucial since they measure the magnitude of default correlation.

More precisely, as far as parameters $\alpha_1$ and $\alpha_2$ are concerned, the absolute bias of the GA estimator is not always smaller than the absolute bias of the CSA estimator. However, its standard error is always smaller and leads to a smaller RMSE for the granularity adjusted estimators. The gain in terms of RMSE is at most 5% (for $\alpha_1, T = 10$ and $N = 50$). For default correlation parameters $\beta_1$ and $\beta_2$, the RMSE gain of the adjusted estimators is much larger: for instance for $T = 50, N = 50$, the gain for $\beta_1$ is $\frac{216 - 142}{216} = 34\%$ and for $\beta_2$ is $\frac{161 - 97}{161} = 40\%$.

The empirical pdf’s displayed in Figures 1 to 8 confirm these results and show that the distributions of the adjusted estimators of $\beta_1, \beta_2$ are much closer to the normal distribution for small $N$; in particular the granularity adjustment provides a skewness correction.
<table>
<thead>
<tr>
<th>N</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_{1N}$</th>
<th>$\hat{\alpha}_2$</th>
<th>$\hat{\alpha}_{2N}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_{1N}$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\beta}_{2N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>B 0.007</td>
<td>−0.007</td>
<td>0.013</td>
<td>−0.012</td>
<td>−0.024</td>
<td>−0.038</td>
<td>0.114</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>SE 0.146</td>
<td>0.139</td>
<td>0.156</td>
<td>0.147</td>
<td>0.300</td>
<td>0.247</td>
<td>0.176</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.146</td>
<td>0.139</td>
<td>0.157</td>
<td>0.148</td>
<td>0.301</td>
<td>0.249</td>
<td>0.209</td>
<td>0.176</td>
</tr>
<tr>
<td>100</td>
<td>B 0.003</td>
<td>−0.003</td>
<td>0.004</td>
<td>−0.006</td>
<td>−0.024</td>
<td>−0.033</td>
<td>0.070</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>SE 0.126</td>
<td>0.123</td>
<td>0.137</td>
<td>0.134</td>
<td>0.252</td>
<td>0.225</td>
<td>0.160</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.126</td>
<td>0.123</td>
<td>0.137</td>
<td>0.134</td>
<td>0.253</td>
<td>0.227</td>
<td>0.175</td>
<td>0.168</td>
</tr>
<tr>
<td>200</td>
<td>B 0.001</td>
<td>−0.002</td>
<td>0.001</td>
<td>−0.003</td>
<td>−0.021</td>
<td>−0.027</td>
<td>0.044</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>SE 0.117</td>
<td>0.116</td>
<td>0.125</td>
<td>0.124</td>
<td>0.241</td>
<td>0.229</td>
<td>0.157</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.117</td>
<td>0.116</td>
<td>0.125</td>
<td>0.124</td>
<td>0.242</td>
<td>0.231</td>
<td>0.163</td>
<td>0.165</td>
</tr>
<tr>
<td>500</td>
<td>B 0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>−0.014</td>
<td>−0.017</td>
<td>0.040</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>SE 0.110</td>
<td>0.110</td>
<td>0.120</td>
<td>0.120</td>
<td>0.240</td>
<td>0.228</td>
<td>0.156</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.110</td>
<td>0.110</td>
<td>0.120</td>
<td>0.120</td>
<td>0.241</td>
<td>0.230</td>
<td>0.162</td>
<td>0.165</td>
</tr>
</tbody>
</table>

Table 1: Bias, standard error and root mean square error of the estimators, $T=10$

<table>
<thead>
<tr>
<th>N</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_{1N}$</th>
<th>$\hat{\alpha}_2$</th>
<th>$\hat{\alpha}_{2N}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_{1N}$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\beta}_{2N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>B 0.006</td>
<td>−0.015</td>
<td>0.012</td>
<td>−0.021</td>
<td>0.027</td>
<td>−0.015</td>
<td>0.108</td>
<td>−0.015</td>
</tr>
<tr>
<td></td>
<td>SE 0.065</td>
<td>0.060</td>
<td>0.069</td>
<td>0.065</td>
<td>0.214</td>
<td>0.141</td>
<td>0.120</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.065</td>
<td>0.062</td>
<td>0.070</td>
<td>0.068</td>
<td>0.216</td>
<td>0.142</td>
<td>0.161</td>
<td>0.097</td>
</tr>
<tr>
<td>100</td>
<td>B 0.002</td>
<td>−0.006</td>
<td>0.006</td>
<td>−0.008</td>
<td>0.022</td>
<td>−0.005</td>
<td>0.052</td>
<td>−0.014</td>
</tr>
<tr>
<td></td>
<td>SE 0.057</td>
<td>0.056</td>
<td>0.061</td>
<td>0.059</td>
<td>0.153</td>
<td>0.117</td>
<td>0.094</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.057</td>
<td>0.056</td>
<td>0.061</td>
<td>0.060</td>
<td>0.155</td>
<td>0.117</td>
<td>0.108</td>
<td>0.085</td>
</tr>
<tr>
<td>200</td>
<td>B 0.002</td>
<td>−0.001</td>
<td>0.002</td>
<td>−0.004</td>
<td>0.017</td>
<td>0.001</td>
<td>0.021</td>
<td>−0.016</td>
</tr>
<tr>
<td></td>
<td>SE 0.052</td>
<td>0.051</td>
<td>0.056</td>
<td>0.056</td>
<td>0.121</td>
<td>0.104</td>
<td>0.084</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.052</td>
<td>0.051</td>
<td>0.056</td>
<td>0.056</td>
<td>0.123</td>
<td>0.104</td>
<td>0.087</td>
<td>0.083</td>
</tr>
<tr>
<td>500</td>
<td>B 0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>−0.001</td>
<td>0.010</td>
<td>0.002</td>
<td>0.000</td>
<td>−0.017</td>
</tr>
<tr>
<td></td>
<td>SE 0.049</td>
<td>0.049</td>
<td>0.053</td>
<td>0.053</td>
<td>0.108</td>
<td>0.101</td>
<td>0.080</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>RMSE 0.049</td>
<td>0.049</td>
<td>0.053</td>
<td>0.053</td>
<td>0.108</td>
<td>0.101</td>
<td>0.080</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Table 2: Bias, standard error and root mean square error of the estimators, $T=50$

6 Concluding Remarks

Qualitative factor models can be used for credit risk analysis [see e.g. Gordy (2003), Hamerle, Liebig, Rosch (2003), Gurtler, Hibbeln, Vohringer (2008)].
stochastic migration models [Gagliardini, Gourieroux (2005), Feng, Gourieroux, Jasiak (2008)], contagion models [Rosch, Waterfeldt (2007)] and more generally for a dynamic analysis of evolutionary distributions, such as an income distribution. Due to the unobservable common factor, the corresponding log-likelihood functions have a complicated integral expression, which renders very difficult the numerical computation of the maximum likelihood estimates. The aim of this paper was to explain how explicit accurate approximations of the ML estimates can be easily derived for static factor models, when the size of the cohorts are either very, or moderately large. This methodology has also been extended to derive the granularity adjustment of the ML estimator, when the underlying factors satisfy a Gaussian dynamic model.

This analysis is a necessary step before analyzing the effect of estimation risk on measures of (credit, or life insurance) portfolio risk [see e.g. Loffler (2003), Hamerle, Rosch (2005), Gourieroux, Zakioan (2009)].


Wilson, T. (1997)b: "Portfolio Credit Risk II", Risk, October

Appendix 1

Asymptotic expansion of the estimator $\hat{\theta}_N$

By definition of estimators $\hat{\theta}$ and $\hat{\theta}_N$, we have:

$$\frac{\partial L_\infty(\hat{\theta})}{\partial \theta} = 0, \quad \frac{\partial L_N(\hat{\theta}_N)}{\partial \theta} = 0.$$

By considering the expansion around $\hat{\theta}$ of the second set of first-order conditions, we get:

$$\frac{\partial L_N(\hat{\theta})}{\partial \theta} \simeq -\frac{\partial^2 L_N(\hat{\theta})}{\partial \theta \partial \theta'} (\hat{\theta}_N - \hat{\theta}).$$

This is equivalent to:

$$\hat{\theta}_N - \hat{\theta} \simeq \left[ -\frac{\partial^2 L_N(\hat{\theta})}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial L_N(\hat{\theta})}{\partial \theta}$$

$$= \left[ -\frac{\partial^2 L_N(\hat{\theta})}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial (L_N - L_\infty)}{\partial \theta} (\hat{\theta})$$

$$\simeq \left[ -\frac{\partial^2 L_\infty(\hat{\theta})}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial (L_N - L_\infty)}{\partial \theta} (\hat{\theta}).$$
Appendix 2
Granularity Adjustments

The granularity adjustments are based on the two following expansions valid for small \( \delta \Omega \):

\[
(\Omega + \delta \Omega)^{-1} = \{\Omega[Id + \Omega^{-1}\delta \Omega]\}^{-1} = (Id + \Omega^{-1}\delta \Omega)^{-1} - \Omega^{-1}\Omega^{-1} + o(\delta \Omega); \quad (A.1)
\]

\[
\log \det(\Omega + \delta \Omega) = \log \det\{\Omega[Id + \Omega^{-1}\delta \Omega]\} = \log \det \Omega + \log \det(Id + \Omega^{-1}\delta \Omega) = \log \det \Omega + Tr(\Omega^{-1}\delta \Omega) + o(\delta \Omega). \quad (A.2)
\]

i) Granularity adjustment of the log-likelihood

We have from (2.7), (2.12), (A.1), (A.2):

\[
L_N(\theta) - L_{\infty}(\theta) = -\frac{1}{2} \sum_{t=1}^{T} Tr[\Omega(\theta)^{-1}\hat{\Sigma}_{Nt}]
\]

\[
+ \frac{1}{2} \sum_{t=12}^{T} [\hat{a}_t - \mu(\theta)]'\Omega(\theta)^{-1}\hat{\Sigma}_{Nt}\Omega(\theta)^{-1}[\hat{a}_t - \mu(\theta)] + o(1/N).
\]

iii) Differential of the granularity adjustment of the log-likelihood

Let us consider the granularity adjustment \( L_{Nt}(\theta) - L_{\infty,t}(\theta) \) of the log-likelihood associated with observation \( t \). Its differential for a change \( \delta \mu \) in the mean is:

\[
\delta[L_{Nt}(\theta) - L_{\infty,t}(\theta)] = -[\hat{a}_t - \mu(\theta)]'\Omega(\theta)^{-1}\hat{\Sigma}_{Nt}\Omega(\theta)^{-1}\delta \mu,
\]
or:

\[ \delta l_{Nt} = -\hat{\epsilon}^t \hat{\Sigma}_{Nt} \hat{\Omega}^{-1} \delta \mu + o(1/N), \quad (A.3) \]

when it is evaluated in the CSA maximum likelihood estimator.

Its differential for a change \( \delta \Omega \) in \( \Omega \) is:

\[
\delta [\mathcal{L}_{Nt}(\theta) - \mathcal{L}_{\infty,t}(\theta)]
= \frac{1}{2} \text{Tr} [\Omega(\theta)^{-1} \delta \Omega \Omega(\theta)^{-1} \hat{\Sigma}_{Nt}]
- [\hat{a}_t - \mu(\theta)]' \Omega(\theta)^{-1} \delta \Omega(\theta)^{-1} \hat{\Sigma}_{Nt} \Omega(\theta)^{-1} [\hat{a}_t - \mu(\theta)].
\]

This implies

\[
\delta l_{Nt} = \text{Tr} \{ \delta \Omega [\frac{1}{2} \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\Omega}^{-1} - \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\epsilon}_t \hat{\epsilon}_t'] \} + o(1/N). \quad (A.4)
\]

We deduce from (A.3), (A.4) that:

\[
\frac{\partial l_{Nt}(\hat{\theta})}{\partial \theta} = - \frac{\partial \mu'(\hat{\theta})}{\partial \theta} \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\epsilon}_t \\
+ \sum_{j=1}^{K_S} \sum_{l=1}^{K_S} \left[ \frac{1}{2} \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\Omega}^{-1} - \hat{\Omega}^{-1} \hat{\Sigma}_{Nt} \hat{\epsilon}_t \hat{\epsilon}_t' \right]_{jl} \frac{\partial \Omega_{jl}(\hat{\theta})}{\partial \theta} + o(1/N) \]
Appendix 3

Inverse and Determinant of Matrix $\Omega$

Let us denote by $\beta_l, l = 1, \ldots, L$, the column vectors of matrix $\beta$ and $\tilde{\beta}_l = \beta_l / (\beta_l' \beta_l)^{1/2}$ the associated rescaled vectors with unit norm. We have:

$$\Omega = \sum_{l=1}^{L} \beta_l \beta_l' + \eta^2 I_d$$

$$= \sum_{l=1}^{L} \tilde{\beta}_l \tilde{\beta}_l' [\eta^2 + \beta_l' \beta_l] + \sum_{l=L+1}^{KS} \tilde{\beta}_l \tilde{\beta}_l' \eta^2,$$

where $\tilde{\beta}_l, l = L + 1, \ldots, KS$ is an orthonormal set of vectors introduced to complete $\tilde{\beta}_l, l = 1, \ldots, L$ into an orthonormal basis.

Therefore, we have:

$$\Omega = \tilde{\beta} \begin{pmatrix} \eta^2 + \beta_1' \beta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \eta^2 \end{pmatrix} \tilde{\beta}' = \tilde{\beta} \begin{pmatrix} \eta^2 + \beta_L' \beta_L & \cdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \eta^2 \end{pmatrix} \tilde{\beta}',$$

where $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_{KS})$ is an orthogonal matrix.

We deduce that:

$$\det \Omega = (\eta^2)^{KS-L} \prod_{l=1}^{L} (\eta^2 + \beta_l' \beta_l),$$

and

$$\Omega^{-1} = \sum_{l=1}^{L} \tilde{\beta}_l \tilde{\beta}_l' \left[ \frac{1}{\eta^2 + \beta_l' \beta_l} - \frac{1}{\eta^2} \right] + \frac{1}{\eta^2} I_d$$

$$= -\sum_{l=1}^{L} \beta_l \beta_l' \frac{1}{\eta^2 (\eta^2 + \beta_l' \beta_l)} + \frac{1}{\eta^2} I_d.$$
Appendix 4

The CSA Maximum Likelihood Estimator

i) Estimation of $\alpha$

Under the restrictions of Section (3.3), the CSA log-likelihood function is given by:

$$\frac{1}{T} L_\infty(\theta) \propto -\frac{1}{2} \log \det \Omega(\beta, \eta^2) - \frac{1}{2T} \sum_{t=1}^T (\hat{a}_t - \alpha)' \Omega(\beta, \eta^2)^{-1} (\hat{a}_t - \alpha).$$

The log-likelihood function is easily concentrated with respect to $\alpha$. Indeed, the generalized least squares estimator of $\alpha$ coincides with the ordinary least squares one, since the Kruskal condition is satisfied [see Kruskal (1968)]. Thus, we get

$$\hat{\alpha} = \frac{1}{T} \sum_{t=1}^T \hat{a}_t \equiv \bar{a}_T. \quad (A.5)$$

ii) Concentrated log-likelihood

We deduce the log-likelihood concentrated with respect to $\alpha$, that is

$$\frac{1}{T} L_\infty^c(\beta, \eta^2) \propto -\frac{1}{2} \log \det \Omega(\beta, \eta^2) - \frac{1}{2T} \sum_{t=1}^T (\hat{a}_t - \hat{\alpha})' \Omega(\beta, \eta^2)^{-1} (\hat{a}_t - \hat{\alpha})$$

$$= -\frac{1}{2} \log \det \Omega(\beta, \eta^2) - \frac{1}{2} Tr \{\Omega(\beta, \eta^2)^{-1} \hat{V}_T\},$$

where $\hat{V}_T = \frac{1}{T} \sum_{t=1}^T (\hat{a}_t - \bar{a}_T)(\hat{a}_t - \bar{a}_T)'$ denotes the sample variance-covariance matrix of the canonical factors and $Tr$ the trace operator. By substituting the explicit expressions of $\det \Omega$ and $\Omega^{-1}$ derived in Appendix 3, we get:
\[ \frac{1}{T} L_\infty^c(\beta, \eta^2) \]

\[ \propto - \frac{KS - L}{2} \log \eta^2 - \frac{1}{2} \sum_{l=1}^{L} \log(\eta^2 + \beta'_l \beta_l) - \frac{1}{2\eta^2} \text{Tr}(\hat{V}_T) \]

\[ + \frac{1}{2} \sum_{l=1}^{L} \beta'_l \hat{V}_T \beta_l - \frac{1}{2} \eta^2[\eta^2 + \beta'_l \beta_l]. \]

**iii) Estimators of \( \beta \) and \( \eta^2 \)**

Let us consider the first-order condition with respect to \( \beta_l \) without taking into account the orthogonality restrictions (3.14). We get:

\[ \frac{1}{T} \frac{\partial L_\infty^c}{\partial \beta_l} = \left\{ - \frac{1}{\eta^2 + \beta'_l \beta_l} - \frac{\beta'_l \hat{V}_T \beta_l}{\eta^2(\eta^2 + \beta'_l \beta_l)^2} \right\} \beta_l + \hat{V}_T \beta_l = 0. \]  

(A.6)

Therefore, the condition \( \frac{1}{T} \frac{\partial L_\infty^c}{\partial \beta_l} = 0 \) implies that \( \hat{V}_T \beta_l \) and \( \beta_l \) are proportional, that is, \( \beta_l \) is an eigenvector of the estimated variance-covariance matrix \( \hat{V}_T \).

Let us now denote by \( \lambda_l, l = 1, \ldots, L \) the eigenvalues of \( \hat{V}_T \), and by \( e_l, l = 1, \ldots, L \) the associated eigenvectors with unit norm and \( \gamma_l = \beta'_l \beta_l \) the square of the norm of the solution. We have:

\[ \beta_l = \gamma_l^{1/2} e_l, \]  

(A.7)

where parameter \( \gamma_l \) satisfies:

\[ - \frac{1}{\eta^2 + \gamma_l} - \frac{\lambda_l \gamma_l}{\eta^2(\eta^2 + \gamma_l)^2} + \frac{\lambda_l}{\eta^2(\eta^2 + \gamma_l)} = 0, \]

by substituting expression (A.7) of \( \beta_l \) in the first-order condition. This provides

\[ \gamma_l = \lambda_l - \eta^2, \]  

(A.8)

and therefore \( \lambda_l = \eta^2 + \beta'_l \beta_l \).
Let us finally concentrate with respect to the optimal $\beta_l$. The concentrated log-likelihood becomes:

$$\frac{1}{T} L_\infty^c(\eta^2)$$

$$\propto - \frac{KS - L}{2} \log \eta^2 - \frac{1}{2} \sum_{l=1}^{L} \log \lambda_l - \frac{1}{2\eta^2} Tr \hat{V}_T$$

$$+ \frac{1}{2} \sum_{l=1}^{L} \frac{(\lambda_l - \eta^2)}{\eta^2}$$

$$= - \frac{KS - L}{2} \log \eta^2 - \frac{1}{2\eta^2}(Tr \hat{V}_T - \sum_{l=1}^{L} \lambda_l) - \frac{1}{2} \sum_{l=1}^{L} \log \lambda_l - \frac{L}{2}.$$

The first-order condition with respect to $\eta^2$ provides:

$$\eta^2 = \frac{Tr \hat{V}_T - \sum_{l=1}^{L} \lambda_l}{KS - L}, \quad (A.9)$$

and the corresponding value of the concentrated log-likelihood is equal (up to an additive constant term) to $- \frac{KS - L}{2} \log(\sum_{l=L+1}^{KS} \lambda_l) - \frac{1}{2} \sum_{l=1}^{L} \log \lambda_l$. This value is maximized, when the $L$ largest eigenvalues are selected. This proves Proposition 2.

iv) The information matrix

Let us introduce the new parameters $\beta^*_l = \beta_l / \eta$ and the notation $B^* = (\beta^*_1, \ldots, \beta^*_L)$. The asymptotic log-likelihood concentrated with respect to $\alpha$ becomes:

$$\lim_{T \to \infty} \frac{1}{T} L_\infty^c \propto - \frac{KS}{2} \log \eta^2 - \frac{1}{2} \sum_{l=1}^{L} \log(1 + \beta_l^* \beta_l^*) - \frac{1}{2\eta^2} Tr \Omega$$

$$+ \frac{1}{2} \sum_{l=1}^{L} \frac{\beta_l^* \Omega \beta_l^*}{\eta^2[1 + \beta_l^* \beta_l^*]},$$

35
where $\Omega/\eta^2 = Id + B^*B^{''}$ evaluated at the true parameter values. The opposite of the Hessian of this asymptotic log-likelihood is easily obtained. We have:

$$-\frac{1}{T} \frac{\partial^2 L^c_\infty}{\partial \beta^l \partial \beta^l} \simeq \frac{(Id - \Omega/\eta^2)\beta^l_i}{1 + \beta^l_i \beta^l_i} + \frac{\beta^l_i \Omega \beta^l_i}{\eta^2 (1 + \beta^l_i \beta^l_i)^2} \beta^l_i.$$

We deduce:

$$-\frac{1}{T} \frac{\partial^2 L^c_\infty}{\partial \eta^2 \partial \beta^l_i} \simeq \frac{1}{\eta^2} \frac{\Omega \beta^l_i}{1 + \beta^l_i \beta^l_i} - \frac{1}{\eta^2} \frac{\beta^l_i \Omega \beta^l_i}{(1 + \beta^l_i \beta^l_i)^2} \beta^l_i$$

$$= \frac{1}{\eta^2} \frac{\beta^l_i}{(1 + \beta^l_i \beta^l_i)},$$

since vectors $\beta^l_i, l = 1, \ldots, L$ are orthogonal.

Similarly, we get:

$$-\frac{1}{T} \frac{\partial^2 L^c_\infty}{\partial \beta^l_i \partial \beta^k_i} \simeq \frac{(Id - \Omega/\eta^2)\beta^l_i \beta^k_i}{1 + \beta^l_i \beta^k_i} + \frac{\beta^l_i \Omega \beta^k_i}{\eta^2 (1 + \beta^l_i \beta^k_i)^2} Id$$

$$- \frac{2(Id - \Omega/\eta^2)\beta^l_i \beta^k_i}{(1 + \beta^l_i \beta^k_i)^2} + \frac{2(\Omega/\eta^2)\beta^l_i \beta^k_i}{(1 + \beta^l_i \beta^k_i)^2}$$

$$= \frac{4}{\eta^2 (1 + \beta^l_i \beta^k_i)} \beta^l_i \beta^k_i$$

$$= \frac{\beta^l_i \beta^k_i}{1 + \beta^l_i \beta^k_i} Id - \frac{B^*B^{''}}{1 + \beta^l_i \beta^k_i} + \frac{2\beta^l_i \beta^k_i}{(1 + \beta^l_i \beta^k_i)^2},$$

and

$$-\frac{1}{T} \frac{\partial^2 L^c_\infty}{\partial \beta^l_i \partial \beta^h_i} \simeq 0, \text{ for } h \neq l.$$
FIGURE 3a: Empirical PDF of the Estimators of $\alpha_1$
True value: $\alpha_1$, $t=10$, $N=200$
Solid line: Granularity-Adjusted Estimator, Dotted line: CSA Estimator

FIGURE 3b: Empirical PDF of the Estimators of $\alpha_2$
True value: $\alpha_2$, $t=10$, $N=200$
Solid line: Granularity-Adjusted Estimator, Dotted line: CSA Estimator

FIGURE 3c: Empirical PDF of the Estimators of $\beta_1$
True value: $\beta_1$, $t=10$, $N=200$
Solid line: Granularity-Adjusted Estimator, Dotted line: CSA Estimator

FIGURE 3d: Empirical PDF of the Estimators of $\beta_2$
True value: $\beta_2$, $t=10$, $N=200$
Solid line: Granularity-Adjusted Estimator, Dotted line: CSA Estimator