Composite Likelihood for Stochastic Migration Model with Unobserved Factor*

Djogbenou, A.¹, Gouriéroux, C.², Jasiak J.¹, and M. Bandehali¹

 $$^{1}\mbox{York}$ University $^{2}\mbox{Toronto}$ University, Toulouse School of Economics and CREST

April 2021

Abstract

We introduce the Maximum Composite Likelihood (MCL) estimation method for the stochastic factor ordered Probit model of credit rating transitions of firms. This model is recommended to banks and financial institutions as part of internal credit risk assessment procedures under the Basel III regulations. However, its exact likelihood function involves a high-dimensional integral, which can be approximated numerically and next this approximation can be maximized. However, the associated estimates of migration risk and corresponding required capital are generally quite sensitive to the quality of this approximation, leading to statistical regulatory arbitrage. The proposed MCL estimator maximizes the composite log-likelihood of the factor ordered Probit model. We present three MCL estimators of different complexity and prove their consistency and asymptotic normality. The performance of these estimators is examined in a simulation study.

Keywords: Migration Model, Credit Rating, Basel III, Composite Likelihood, Factor Model, Large Panel, Statistical Regulatory Arbitrage.

^{*}The authors gratefully acknowledge financial support of the chair of the Authority of Prudential Control and Resolution (ACPR)/Risk Foundation: Regulation and Systemic Risks, the ECR DYSMOIA and the Natural Sciences and Engineering Council of Canada (NSERC).

1 Introduction

Under the "internal-ratings-based" (IRB) approach advocated in the Basel II and III regulation, banks use their internal risk rating systems to estimate the risk exposures, credit rating migration probabilities and the probability of default (PD) in order to evaluate their regulatory capital requirements [See e.g. Basel Committee on Banking Supervision (2004, 2009), Hull (2012)]. Under Pillar II, financial institutions must also conduct stress tests to determine the proper level of capital needed to absorb losses in worsening economic conditions. For these reasons, banks perform their own credit rating migration analysis in order to monitor the changes in borrowers' credit quality and to predict borrowers' potential default in a volatile economic environment. This analysis concerns the "internal" or "in-house" established credit rating histories of borrowers that can be classified into various numbers of credit quality categories determined independently of the ratings publicly provided by the rating agencies such as the Moody's. The internal credit rating analysis concerns the historical probabilities of default and migration probabilities. It differs from the analysis of their risk-neutral counterparts, which underlies the pricing of credit derivatives, such as credit default swaps (CDS), or derivatives written on iTraxx [see, e.g. Duffie, Eckner, Horel, Saita (2009), Azizpour, Giesecke, Schwenkler (2018) in continuous time, Gouriéroux, Monfort, Polimenis (2006) in discrete time, Gouriéroux, Monfort, Mouabbi, Renne (2021) for joint historical and risk-neutral analysis. The internal ratings are used for pricing the portfolios of credits to a large number of small and medium-size firms whose assets are not traded on the markets. Even for large firms, the historical and risk-neutral probabilities of default can differ significantly. Therefore, the analysis of internal ratings is consistent with prudential banking supervision and aims at avoiding a pure mark-to-market pricing of risk.²

The credit rating migration analysis concerns the changes [i.e. upgrades or downgrades] of borrowers' credit quality over time with respect to their previous ratings [Altman, Saunders (2004)]. These are recorded in the form of monthly or quarterly time series of credit migration matrices comprising qualitative ratings of firm, ranked from the low risk category A to the most risky rating D, of default. The ordered Probit model for credit ratings arises as a natural specification, which has been extended to the Asymptotic Stochastic Factor Model (ASFR), by Vasicek (1991) [see also Vasicek (2015), Nickell, Perraudin, Varotto (2001)] and recommended under the Basel III regulatory measures. The ASFR is a stochastic factor probit model of default with an independent and identically distributed common random unobserved factor, capturing the systematic risk effect. The factor is assumed to drive the parameters of a latent quantitative score function in the model, which determines the observed qualitative ratings. Due to the presence of the unobserved common factor, the observed rating histories are cross-sectionally dependent, which can explain default correla-

¹Publicly available credit ratings of large obligors are available from the rating agencies such as the Moody's, Standard and Poor, and Fitch.

²There is often a confusion about the notions of historical and risk-neutral risks. For example, Moody's Analytics provides "EDF" estimates of the historical probability of default by considering default frequency of firms at the same distance-to-default (DD). However, the notion of DD is risk-neutral.

tion. Koopman, Lucas, Monteiro (2008), Feng, Gouriéroux, Jasiak (2008), Creal, Koopman, Lucas (2012), and Creal, Schwaab, Koopman, Lucas (2014) extended this setup to multiple credit rating categories and a serially correlated factors, for predicting the future credit ratings of firms.

The estimation of the ordered Probit model with a latent factor is challenging. In order to derive the joint density of observed ratings, the history of the latent factor has to be integrated out. Therefore, the exact likelihood function based on the joint density of rating histories involves an integral of high dimension that increases with the number of observations. Due to the presence of these integrals, the exact maximum likelihood is commonly replaced by an approximation. There exist various approximation methods, most of which involve a set of arbitrary control parameters, which have a significant impact on the associated required capital. These are, for example, the discretization step [Farmer (2021)], smoothing parameters, penalization rate in neural networks, etc. The effect of the statistical approximation and optimization method can go as far as to partly circumvent the need for keeping an internal capital reserve. Therefore, it is called a "statistical regulatory arbitrage". This explains why these approximations are generally not validated by the supervisory authorities who are regularly auditing the internal databases and estimation techniques.³ So far, the banking supervisory authority has validated standardized approximation methods, such as the granularity adjusted approximation, corresponding to large cross-sectional asymptotics [see, Gagliardini, Gouriéroux (2015) for general discussion] and the Simulated Maximum Likelihood method with a large number of simulations.

This paper introduces a set of Maximum Composite Likelihood (MCL) estimators for the stochastic factor ordered Probit model, as simple alternative methods providing consistent and normally distributed estimators. The MCL estimation method has been widely used in the statistical literature to handle complex likelihood functions [for example, see, e.g. Lindsay (1988), Varian (2008), Varian, Reid, Firth (2011) and Gouriéroux, Monfort (2018)]. The composite likelihood functions are obtained by multiplying a collection of component likelihoods and are known to provide consistent parameter estimators (Varian, Reid, Firth, 2011). In this paper, the MCL estimator maximizes an objective function based on the exact likelihood function of the factor ordered Probit model. We propose three MCL estimators of different complexity and prove their consistency and asymptotic normality. The performance of the estimators is examined in a simulation study.

This paper is organized as follows. Section 2 describes the ordered probit model for credit rating transitions. Section 3 describes the stochastic factor ordered Probit model. Section 4 introduces the composite maximum likelihood estimators and derives their consistency and asymptotic normality. The simulation results are presented in Section 5. Section 6 concludes the paper. Proofs, simulation details and additional simulation results are gathered in Appendices.

³The Bayesian methods that are sensitive to the choice of the prior, are systematically not validated by the banking supervisory authority [see, e.g. Duffie, Eckner, Horel, Saita (2009) for a Bayesian estimation method].

2 The Stochastic Factor Ordered-Probit Model

In this section, we introduce the stochastic factor ordered probit model and its state space representation. Next, we derive the expression of the complete likelihood function to highlight the presence of multiple integrals of large dimension.

2.1 The State-Space Representation

Let $y_{i,t}^*$ and $y_{i,t}$ denote the (credit) score and rating of firm i, i = 1, ..., N at time t, t = 1, ..., T. The latent continuous quantitative score (y_{it}^*) determines the individual qualitative rating y_{it} . More precisely, variables y_{it} define the qualitative individual histories of credit ratings with K rating categories : k = 1, ..., K. The score is discretized in order to obtain the individual qualitative ratings. Therefore, a rating is determined by:

$$y_{i,t} = k$$
, if and only if $c_k \le y_{i,t}^* < c_{k+1}, \quad k = 1, ..., K$, (2.1)

where $c_1 < \cdots < c_{K+1}$ are thresholds. Relation (2.1) shows how the observable endogenous credit rating $(y_{i,t})$ is linked to the latent score function $(y_{i,t}^*)$. By convention, we have $c_1 = -\infty$ and $c_{K+1} = +\infty$. Moreover, relation (2.1) defines the measurement equation of the state space representation of the model.

The conditional distribution of the quantitative scores given the past depends on the common latent factor f^4 and on the past individual ratings $y_{i,t-1}$, such that:

$$y_{i,t}^* = \delta_l + \beta_l f_t + \sigma_l u_{i,t}, \quad i = 1, ..., n, \quad \text{if } y_{i,t-1} = l, l = 1, ..., K.$$
 (2.2)

The multivariate, continuous, latent processes y_{it}^* , are generated by individual level effects, (δ_l) , volatility effects (σ_l) , $\sigma_l > 0$, factor effects where the components of β_l define the factor sensitivities. When coefficient β is large (small, resp.), the effect of systematic risk carried through the factor is strong (weak, resp.). While the idiosyncratic risks $(u_{i,t})$ can be diversified, the systematic risk (f_t) cannot be diversified. Thus the presence of systematic risk generates risk dependence in the model. The following autoregressive model of order 1 (AR(1)) represents the common factor dynamics:

$$f_t = \rho f_{t-1} + \sqrt{1 - \rho^2} \eta_t, |\rho| < 1, \tag{2.3}$$

where η_t defines the shock to the common factor. The system of equations (2.2)-(2.3) defines the state equations of the state-space model. Let us introduce the following assumptions:

Assumption A.1: The errors $u_{i,t}$, η_t , i = 1, ..., n, t = 1, ..., T are independent, standard normal variables.

⁴Alternatively, a multidimensional factor can be considered to distinguish between the dynamic patterns of migrations of firms with good and poor credit quality, respectively.

The independence assumption allows for performing impulse response analysis by shocking separately the idiosyncratic and systematic innovations. The assumption of identical distribution and the fact that coefficients in (3.2) are independent of the firm implies that we consider a homogeneous set of firms, obtained by crossing the country, the industrial sector and the size, in conformity with the current regulation.

Assumption A.2: The factor process (f_t) is the strongly stationary solution of autoregressive equation (2.3).

As the processes (f_t) , $(u_{i,t})$, $i=1,\ldots,n$ are independent and strictly stationary, it follows that the joint n-dimensional process $y_t^* = \left(y_{1,t}^*, \ldots, y_{n,t}^*\right)'$ is also strictly stationary, and so is its state discretized version $y_t = (y_{1,t}, \ldots, y_{n,t})'$. However, the individual components $(y_{i,t}^*)$, $i=1,\ldots,n$ are not independent due to the effect of the common factor f_t . Note also that the error variance in equation (2.3) has been set equal to $1-\rho^2$. This implies that factor f_t is marginally distributed with mean 0 and variance 1. This constraint is imposed on the marginal distribution of the factor for identification of parameters δ_l , β_l , $l=1,\ldots,K$.

Assumption A.3: The variables $(y_{i,t})$ are observed and variables $(y_{i,t}^*)$, (f_t) are latent.

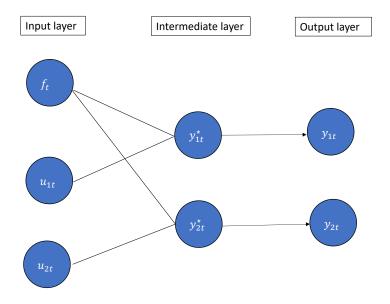
This assumption explains the state space interpretation of the model. Moreover, a neural network framework is also available, if we distinguish three layers: the deep layer generates the factors values and the idiosyncratic errors; the intermediate layer computes the latent scores; the output layer provides the observed rating. By analogy with the graphical neural network representation, we obtain the following scheme displayed in Figure 1 (given with n = 2 and K = 3).

In practice, the underlying quantitative scores are computed by a credit institution and each individual (i.e. firm) can request the records of its own score history. However, the complete score database is, in general, proprietary and the information on the quantitative scores is not available to an outsider econometrician/data scientist. This is the rationale for introducing Assumption A3.

The above state space specification differs from the autoregressive probit panel model studied in Tuzcuoglu (2019), who considers idiosyncratic correlations, but does not include the systematic factor component required in the Basel regulation for stress tests.

The factor f_t is assumed unobserved for the following two reasons: First to create the cross-sectional correlation between individual risks. Second, to get a complete dynamic model that can be used to predict the defaults in the future. A bias would result from directly replacing factor f_t by an observed proxy \hat{f}_t , such as a combination of market volatility index VIX, a consumer sentiment index, consumption growth, a business cycle indicator [see e.g. Berndt, Douglas, Duffie, Fergusson (2018), Azizpour, Giesecke, Schwenkler (2018)], or the slope of the yield curve. Moreover, with observed factors, the predictions could not be performed without introducing an additional model for all components of \hat{f}_t .

Figure 1: Graphical Representation in Neural Network when n=2 and K=3



2.2 The Complete Likelihood Function

In order to derive the joint density of observations y_{it} , i = 1, ..., n, t = 1, ..., T, the unobserved factor history has to be integrated out. As a consequence, observations $y_{i,t}$ are cross-sectionally dependent and serially dependent with a non-Markovian serial dependence. More precisely, the stochastic migration probabilities between dates t - 1 and t, conditional on f_t , are given by:

$$p_{l,k,t} = p_{kl}(f_t; \theta) = P[y_{i,t} = k | y_{i,t-1} = l, f_t]$$

$$= P[c_k \le y_{i,t}^* < c_{k+1} | y_{i,t-1} = l]$$

$$= \Phi\left(\frac{c_{k+1} - \beta_l f_t - \delta_l}{\sigma_l}\right) - \Phi\left(\frac{c_k - \beta_l f_t - \delta_l}{\sigma_l}\right), l, k = 1, ..., K, t = 2, ..., T,$$
(2.4)

where Φ denotes the cumulative distribution function (c.d.f.) of the standard normal. Thus each row of the transition matrix conditional of (f_t) contains an ordered polytomous model with a common explanatory factor f_t . When factor f_t is unobserved stochastic and serially correlated, as in (2.3), the transition matrices are stochastic and serially dependent.

Let us now define the log-likelihood function of the stochastic migration model. The vector θ includes the parameters of the state space model, which are parameters δ_l , β_l , σ_l , $l = 1, \ldots, K$ in the quantitative score, and parameters c_k , $k = 2, \ldots, K$ defining the states. As

the conditional migration matrices are functions of parameter vector θ as well as of the common factor values (f_t) , the likelihood function conditional on (F) is:

$$L_T(Y|F, y_1; \theta) = \prod_{t=2}^{T} \prod_{k=1}^{K} \prod_{l=1}^{K} (p_{l,k}(f_t; \theta))^{n_{l,k,t}},$$
(2.5)

where $n_{l,k,t}$ denotes the number of firms which migrate from l to k between t-1 and t, $Y = (y_{i,t})$ for i = 1, ..., n and t = 2, ..., T, $F = (f_t)$, t = 2, ..., T and $y_1 = (y_{1,1}, ..., y_{n,1})'$.

Since the factor history is not observed, the distribution of factor values $(f_1, ..., f_T)$ has to be integrated out and the log-likelihood function, given the initial value y_1 only, is:

$$\ell(Y|y_1;\theta,\rho) = \log \int \dots \int \prod_{t=2}^{T} \prod_{k=1}^{K} \prod_{l=1}^{K} \left[(p_{l,k}(f_t;\theta))^{n_{l,k,t}} \psi(f_2,...,f_T;\rho) \right] df_2...df_T, \tag{2.6}$$

where ψ refers to the joint probability distribution function (p.d.f.) of factor values. The above log-likelihood function contains a multivariate integral. The dimension of this integral is of order T, as there is a common factor value for each transition at time t. Therefore the exact computation of this likelihood is impossible and its approximation is often not sufficiently robust (see, Section 1). The MCL estimators are convenient alternatives for complicated nonlinear dynamic state-space models allowing for circumventing the high-dimensional integral.

3 Composite Likelihood for Migration Model with Unobserved AR(1) Factor

3.1 Transition Probabilities

The process of transition matrices $\{P_t, t = 1, ..., T\}$ has component matrices $P_t = (p_{kl,t})$, which provide the probabilities of transitions from state l to state k between times t-1 and t given f_t . From (2.4), it follows that the elements of matrix P_t are:

$$p_{kl,t} = p_{kl}(f_t; \theta) = \mathbb{P}[y_{i,t} = k | y_{i,t-1} = l, f_t] = \Phi\left(\frac{c_{k+1} - \beta_l f_t - \delta_l}{\sigma_l}\right) - \Phi\left(\frac{c_k - \beta_l f_t - \delta_l}{\sigma_l}\right), \ k, l = 1, ..., K.$$

Let us now compute the product of two successive transition matrices $P_t^{(2)} = P_t P_{t-1}$ to

obtain the probabilities of transition at horizon 2 from state l to k between times t-2 and t given (f_t) . The elements of matrix $P_t^{(2)}$ depend on f_t , f_{t-1} and are given by:

$$p_{kl,t}^{(2)} = p_{kl}(f_t, f_{t-1}; \theta) = \mathbb{P}[y_{i,t} = k | y_{i,t-2} = l, f_t, f_{t-1}] = \sum_{j=1}^{K} [p_{kj}(f_t, \theta) p_{jl}(f_{t-1}, \theta)].$$
 (3.1)

They can be computed from the elements of matrices P_t and P_{t-1} . Let us denote by P and P(2) the expectations of matrices P_t and $P_{t-1}^{(2)}$ with respect to the common factor:

$$P = E(P_t), \quad P(2) = E(P_{t-1}^{(2)}) = E(P_t P_{t-1}).$$
 (3.2)

The elements of matrix P:

$$P = [p_{kl}] = [p_{kl}(\theta)] = E_{f_t}[p_{kl}(f_t, \theta)],$$

are obtained by integrating out the unobserved factor factor value f_t .

Lemma 1 Under Assumptions A1, A2 and A3, we have

$$p_{kl}(\theta) = \mathbb{P}[y_{i,t} = k | y_{i,t-1} = l] = \mathbb{P}[c_k < y_{i,t}^* < c_{k+1} | y_{i,t-1} = l] = \Phi\left(\frac{c_{k+1} - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2}}\right) - \Phi\left(\frac{c_k - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2}}\right).$$

Proof. See Appendix A.1.

The elements of matrix P(2) are:

$$P(2) = [p_{kl}^{(2)}] = [p_{kl}(2; \theta, \rho)] = E_{f_t, f_{t-1}} \left[\sum_{j=1}^{K} p_{kj}(f_t, \theta) p_{j,l}(f_{t-1}, \theta) \right],$$
or, $p_{kl}(2; \theta, \rho) = \mathbb{P}[c_k < y_{i,t}^* < c_{k+1} | y_{i,t-2} = l].$

Lemma 2 Under Assumptions A1, A2 and A3, we have

$$p_{kl}(2;\theta,\rho) = \int \sum_{j=1}^{K} \left[\left[\Phi \left(\frac{c_{k+1} - \delta_j - \beta_j \rho f}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}} \right) - \Phi \left(\frac{c_k - \delta_j - \beta_j \rho f}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}} \right) \right] * \left[\Phi \left(\frac{c_{j+1} - \delta_l - \beta_l f}{\sigma_l} \right) - \Phi \left(\frac{c_j - \delta_l - \beta_l f}{\sigma_l} \right) \right] * \phi(f) df,$$

where ϕ is the p.d.f. of the standard normal.

Proof. See Appendix A.2.

The transitions at horizon 2 involve one-dimensional integrals only, which are easy to compute numerically.

3.2 Conditional Composite Likelihood Functions

This section presents the conditional composite likelihood function for the migration model with an unobserved AR(1) factor. The composite likelihoods [see Cox, Reid (2004), Varian, Reid, Firth (2011)], are often based on misspecified likelihoods, which are easier to calculate. In our framework, the composite likelihoods are constructed from the exact expected migration probabilities at horizons 1 and 2 to reduce the dimension of the integrals.

i) The Conditional Composite Log-Likelihood at Lag 1

The conditional composite log-likelihood function at lag 1, called CL(1) and denoted by $L_{cc}(\theta)$, is defined as:

$$L_{cc}(\theta) = \sum_{i=1}^{n} \sum_{t=2}^{T} log(l(y_{i,t}|y_{i,t-1};\theta))$$

$$= \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{l=1}^{K} [\mathbb{1}(y_{i,t} = k, y_{i,t-1} = l) log(p_{kl}(\theta))]$$

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} [n_{kl} log(p_{kl}(\theta))], \qquad (3.3)$$

where $n_{kl} = \sum_{i=1}^{n} \sum_{t=2}^{T} \mathbb{1}(y_{i,t} = k, y_{i,t-1} = l) = \sum_{t=2}^{T} n_{kl,t}$ is the number of transitions from l to k in one step over the period. The log-likelihood L_{cc} is calculated as if the observed ratings $(y_{i,t})$, i = 1, ..., n, were independent across the individuals, while in reality they are linked by the common factor. Moreover, L_{cc} considers the rating processes $(y_{i,t})$, i = 1, ..., n as if these were components of a Markov chain with transition matrix P, while $(y_{i,t})$, i = 1, ..., n are not Markov due to the factor integration, which increases the memory of the process. Therefore, the CL(1) is a quasi (pseudo) log-likelihood. The composite log-likelihood CL(1) depends on parameter vector θ only, and cannot be used to estimate the factor dynamic, i.e, the autoregressive coefficient ρ . For that purpose, it is necessary to increase the lag.

ii) The Conditional Composite Log-Likelihood at Lag (2)

The conditional composite log-likelihood at lag (2), called CL(2) and denoted by $L_{cc,2}(\theta,\rho)$,

is:

$$L_{cc,2}(\theta,\rho) = \sum_{i=1}^{n} \sum_{t=2}^{T} \left[log(l(y_{i,t}|y_{i,t-2};\theta,\rho)) \right]$$

$$= \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{l=1}^{K} \left[\mathbb{1}(y_{i,t}=k) \mathbb{1}(y_{i,t-2}=l) \ log(p_{kl}(2;\theta,\rho)) \right]$$

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \left[n_{kl}^{(2)} \ log \ p_{kl}(2;\theta,\rho) \right], \tag{3.4}$$

where $n_{kl}^{(2)} = \sum_{t=2}^{T} n_{kl,t}^{(2)}$ is the number of transitions from state l to k in two steps, with

$$n_{kl,t}^{(2)} = \sum_{j=1}^{K} (\hat{p}_{kj,t} \ n_{jl,t-1}).$$

The composite log-likelihood function $L_{cc,2}(\theta, \rho)$ is computed from the density of $(y_{i,t})$ conditional on $(y_{i,t-2})$ as if the rating histories $(y_{i,t})$ were cross-sectionally independent from one another and $(y_{i,t-2})$ were containing all information about the past. Therefore, the CL(2) is a quasi (pseudo) log-likelihood too.

An important difference between L_{cc} and $L_{cc,2}$ is the set of identifiable parameters. As mentioned above, we can expect to identify θ from L_{cc} , but we cannot identify parameter ρ characterizing the cross-sectional dependence. $L_{cc,2}$ provides additional information that is sufficient to identify ρ .

iii) The Conditional Composite Likelihood up to Lag 2

The conditional composite log-likelihood up to lag 2, CL(1,2), is defined by summing up the previous composite log-likelihoods at lags 1 and 2:

$$L_c(\theta, \rho) = L_{cc}(\theta) + L_{cc,2}(\theta, \rho). \tag{3.5}$$

This artificial objective function cannot be interpreted as a quasi likelihood.

iv) The Conditional Log-likelihood

As shown in Section 2.2, the complete log-likelihood has a complicated expression including a high-dimensional integral, with the dimension increasing with T. From the granularity theory [Gagliardini, Gouriéroux (2014, 2015)], it follows that when n tends to infinity, an estimator of θ , asymptotically equivalent to the ML estimator can be obtained from the log-likelihood conditional on the factor path by maximizing this conditional log-likelihood with respect to both parameter θ and the factor path. This log-likelihood conditional on

 (f_t) is:

$$L(\theta, f_1, ..., f_T) = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{l=1}^{K} [\mathbb{1}(y_{i,t} = k, y_{i,t-1} = l) \log p_{kl}(f_t; \theta)]$$

$$= \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{l=1}^{K} [n_{kl,t} \log p_{kl}(f_t; \theta)].$$
(3.6)

It resembles the composite log-likelihood L_{cc} except that in the composite log-likelihood $p_{kl}(\theta)$ has been made independent of f_t by marginalizing. Since it is maximized with respect to θ, f_1, \ldots, f_T , it provides not only an estimator of θ , but also an approximation \hat{f}_t of the factor values. In the second step, an estimator of ρ is obtained by regressing \hat{f}_t on $\hat{f}_{t-1}, t = 2, \ldots, T$.

v) Loss of information in the CL approaches

Intuitively, under an identification restriction, the composite log-likelihood estimator at lag 1, $\tilde{\theta}_T = Argmax_{\theta}L_{cc}(\theta)$, is a consistent estimator of the true parameter value. However, as it does not take into account the serial dependence due to the factor, a loss of information results and the estimator is not asymptotically efficient. The estimator of θ obtained by maximizing the conditional log-likelihood (3.6) captures that serial dependence through the "nuisance" parameters $f_1, ..., f_T$ and achieves the asymptotic efficiency, according to the granularity theory.

By considering the conditional composite log-likelihood up to lag 2 that involves the serial dependence parameter ρ , we expect to partly reduce the lack of efficiency for θ under the CL(1) approach.

In the next subsection, the identification constraints, and the order and rank conditions for each of the conditional composite log-likelihoods are discussed.

3.3 Identification

The parameters to be identified and their respective numbers are as follows:

$$c_k, k = 2, ..., K, number : K - 1,$$

 $\delta_k, k = 1, ..., K, number : K,$
 $\beta_k, k = 1, ..., K, number : K,$
 $\sigma_k, k = 1, ..., K, number : K,$
 $\rho, number : 1.$

The total number of independent parameters to identify is 4K-2. The negative two is due to the score $y_{i,t}^*$ being defined up to an increasing function. As we have supposed that it was

a linear function of factor f_t , the score $y_{i,t}^*$ is defined up to a linear affine increasing function. The intercept and slope of that linear function are not identifiable.

3.3.1 Order Conditions

In this subsection, the order conditions for each conditional composite log-likelihood are discussed.

i) Identification under CL(1):

The identifying functions are the reduced form parameters involved in CL(2), i.e. the elements $p_{kl}(\theta)$ of matrix P. There are K(K-1) of these elements that are linearly independent due to the unit mass restriction on each column. Hence, the order condition is:

$$K(K-1) \geqslant 4K-1 \iff K^2 - 5K + 1 \geqslant 0,$$

by taking into account the absence of parameter ρ in the objective function. This order condition is satisfied for $K \geq 5$.

ii) Identification under CL(2):

The identifying functions are determined by observing that the factor f varies within the integral expression of $p_{k,l}(2;\theta,\rho)$ (see Lemma 2). These identifying functions and their respective numbers are as follows:

(1)
$$\frac{c_k - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)}}; \quad number : \quad K(K - 1);$$

(2)
$$\frac{\epsilon \beta_l \rho}{\sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)}}; \quad number : K;$$

(3)
$$\frac{c_k - \delta_l}{\sigma_l}$$
; number: $K(K-1)$;

(4)
$$\frac{\epsilon \beta_l}{\sigma_l}$$
; number: K ,

where $\epsilon = \pm 1$ is an unknown sign, since the distribution of f is symmetric. This implies that the integral expression in Lemma 2 is also valid with f replaced by -f. There is only one such invariance property and therefore the sign ϵ is equal for all l. The total number of identifying functions of parameters is $2K(K-1) + 2K = 2K^2$. Hence, the order condition is:

$$2K^{2} \geqslant 4K - 2 \iff K^{2} - 2K + 1 \geqslant 0$$
$$\iff (K - 1)^{2} \geqslant 0.$$

The order condition holds for any K.

iii) Identification under CL(1,2):

The total number of functions available is equal to the sum of functions available for each component of the total composite log-likelihood. Therefore, the order condition is:

$$3K^2 - K \geqslant 4K - 2.$$

The order condition is satisfied for any K.

3.3.2 Rank Conditions

Proposition 1 Under the CL(1) log-likelihood function and the identifying constraints $c_2 = 0, \gamma_1 = 1$, we can identify the thresholds c_k , k = 2, ..., K, the intercepts δ_l , l = 1, ..., K, and the $\gamma_l = \sqrt{\beta_l^2 + \sigma_l^2}$, l = 2, ..., K.

Proof. See Appendix B.1.

Proposition 2 Under the CL(2) composite log-likelihood function and the identifying constraints $c_2 = 0, \gamma_1 = \sqrt{\sigma_1^2 + \beta_1^2 (1 - \rho^2)} = 1$, all parameters are identified up to the common sign ϵ for β_l , l = 1, ..., K.

Proof. See Appendix B.2.

In order to identify the unknown sign ϵ , an additional constraint needs to be introduced such as:

$$\beta_1 > 0$$
.

The unknown sign ϵ is a problem of global identification and not of local identification. Hence, when the asymptotic properties of the estimators are derived (see Section 4), this inequality constraint has to be taken into account to obtain the consistency of the estimator. It has no effect on asymptotic normality. The asymptotic properties of the composite log-likelihood estimators are discussed in the next Section.

4 Asymptotic Properties of Composite Log-likelihood Estimators

4.1 The Asymptotics

In a panel data framework, the asymptotic analysis can be performed with respect to the cross-sectional dimension n and time T that can tend to infinity as follows:

- (i) $n \to \infty$, T fixed: short panel asymptotics;
- (ii) n fixed, $T \to \infty$: time series asymptotics;
- (iii) Both $n, T \to \infty$: double asymptotics.

The double asymptotics in case (iii) has been recently developed for applications to big data [Gagliardini, Gouriéroux (2014, 2015), Bonhomme, Jochmans, Robin (2017)]. It corresponds to a long panel of high dimensional time series.

In the migration model with an unobserved factor, the asymptotic analysis existing in the literature concerns the granularity adjusted version of the (complete) maximum likelihood method, i.e. the estimation of θ , f_1 , ..., f_T based on the log-likelihood (3.4) [see Gagliardini, Gouriéroux (2014, 2015)]. Let us denote the maximizers of equation (3.4) by $\hat{f}_{n,t}$, t = 1, ..., T and $\hat{\theta}_{n,T}$, and the autoregressive coefficient estimator obtained by regressing $\hat{f}_{n,t}$ on $\hat{f}_{n,t-1}$, t = 1, ..., T by $\hat{\rho}_{n,T}$. Then, we obtain the following results:

- (i) If $n \to \infty$, $T \to \infty$,
- **a.** $\hat{\theta}_{n,T}$ is consistent of θ , asymptotically normal and converges at speed $\frac{1}{\sqrt{nT}}$.
- **b.** $\hat{f}_{n,t}$ is consistent of f_t , asymptotically normal and converges at speed $\frac{1}{\sqrt{n}}$, for any t = 1, ..., T (but the convergence is not necessarily uniform in t).
- **c.** $\hat{\rho}_{n,T}$ is consistent of ρ , asymptotically normal and converges at speed $\frac{1}{\sqrt{T}}$.

Depending on the setup, other asymptotic results can be considered. For example:

- (ii) If $n \to \infty$, T is fixed,
- **a.** $\hat{\theta}_{n,T}$ converges to a stochastic limit, is consistent of θ , asymptotically normal and converges at speed $\frac{1}{\sqrt{n}}$.
- **b.** $\hat{f}_{n,t}$ is consistent of f_t , asymptotically normal and converges at speed $\frac{1}{\sqrt{n}}$, for any t.
- **c.** $\hat{\rho}_{n,T}$ is inconsistent.
- (iii) If n is fixed, $T \to \infty$, neither $\hat{\theta}_{n,T}$, nor $\hat{f}_{n,t}$, $t = 1, \ldots, T$, nor $\hat{\rho}_{n,T}$ are consistent.

In this latter case, it is easy to see that the regression used to estimate ρ involves non negligible measurement errors $\hat{f}_{n,t} - f_t, t = 1, \dots, T$ in the regressors, which explains the lack of consistency.

The asymptotic properties of the maximum conditional composite likelihood estimators are much easier to derive than the asymptotic properties of the complete ML estimator. Indeed, the composite log-likelihood functions are finite sums of products of summary statistics and functions of parameters. This simplifies the proof of uniform convergence with respect to the parameters. The next section examines the asymptotics (i) -(iii) and describes the properties of the conditional composite maximum likelihood estimators.

4.2 Consistency

This section examines the consistency of the maximum conditional composite likelihood estimators of the identifiable parameters. To prove the consistency, we need the following additional assumption:

Assumption A4

- i) The parameter set of (θ, ρ) is compact, and strictly included in the set $\sigma_l > 0, \forall l, |\rho| < 1$.
- ii) The model is well-specified, i.e. the true value (θ_0, ρ_0) is in the interior of the parameter set.

The condition $\sigma_l > 0$, $\forall l$, ensures that the transition probabilities $p_{kl}(f_t; \theta)$ [resp. $p_{kl}(\theta), p_{kl}(2; \theta, \rho)$] are infinitely continuously differentiable with respect to f_t and θ (resp. with respect to θ, ρ).

(i) Short panel asymptotics: $n \to \infty, T$ fixed.

As for the granularity-approximated complete ML estimators, we cannot expect the composite ML estimators to be consistent, for T fixed. This is a consequence of the cross-sectional dependence due to the common systematic factor f_t . To clarify this point, let us assume T=2 and consider the maximum conditional composite likelihood estimator based on CL(1). For T=2, the conditional composite log-likelihood is:

$$L_{cc,n}(c,\delta,\gamma) = \sum_{k=1}^{K} \sum_{l=1}^{K} [n_{kl,2} \log p_{kl}(c,\delta,\gamma)],$$

with $n_{kl,2}$ is $n_{kl,t}$ for t=2, and $\theta=(c,\delta,\gamma)$ is the identifiable parameter, where c,δ,γ are the vectors with elements c_k,δ_l,γ_l , respectively, and take into account the identification restriction in Proposition 1, that are $c_2=0,\gamma_1=1$. By Assumptions A.1, A.2 and the fact that the rating indicators are nonnegative and bounded, we can apply the Strong Law of

Large Numbers. The normalized log-likelihood tends a.s. to:

$$\lim_{n \to \infty} \text{a.s.} \frac{1}{n} L_{cc,n}(c, \delta, \gamma) = \lim_{n \to \infty} \text{a.s.} \sum_{l=1}^{K} \left(\frac{n_l}{n} \sum_{k=1}^{K} \left[\frac{n_{kl,2}}{n_l} \log p_{kl}(c, \delta, \gamma) \right] \right),$$

where $\frac{n_l}{n}$ is the frequency of individuals in category l. We have:

$$\lim_{n \to \infty} \text{a.s.} \frac{n_l}{n} = \mathbb{P}[c_l < y_{i,2}^* < c_{l+1}] = p_l(\theta_0),$$

where $p_l(\theta_0)$ is the stationary probability of being in rating category l evaluated at the true parameter value $\theta_0 = (c_0, \delta_0, \gamma_0)$. Therefore, we get:

$$\lim_{n \to \infty} \text{a.s.} \frac{1}{n} L_{cc,n}(c, \delta, \gamma) \simeq \sum_{l=1}^{K} \left[p_l(\theta_0) \left[\sum_{k=1}^{K} (\lim_{n \to \infty} \text{a.s.} \hat{p}_{kl,2}) \log p_{kl}(c, \delta, \gamma) \right] \right]$$
$$= \sum_{l=1}^{K} \left[p_l(\theta_0) \left[\sum_{k=1}^{K} p_{kl,1}(\theta_0, f_2) \log p_{kl}(c, \delta, \gamma) \right] \right] \equiv L_{cc,\infty}.$$

 $\frac{1}{n}L_{cc,n}(c,\delta,\gamma)$ and its limit $L_{cc,\infty}(c,\delta,\gamma)$ are continuous functions of the identifiable parameters. Then, they admit at least a maximum on the parameter set by Assumption A4 ii). Let us denote by θ_0^* , the pseudo-true value, i.e. a solution of the asymptotic optimization problem, we have:

$$\theta_0^* = (c_0^*, \delta_0^*, \gamma_0^*) = \underset{c, \delta, \gamma}{\operatorname{argmax}} \sum_{l=1}^{L} \left[p_l(\theta_0) \left[\sum_{k=1}^{K} p_{kl,1}(\theta_0, f_2) \log p_{kl}(c, \delta, \gamma) \right] \right].$$

By applying the Jennrich inequality [Jennrich (1969)], if the pseudo true value is unique, there exists a sequence of maximum conditional composite likelihood estimators $\hat{\theta}_n = (\hat{c}_n, \hat{\delta}_n, \hat{\gamma}_n) = Argmax_{c,\delta,\gamma} \frac{1}{n} L_{cc,n}(c,\delta,\gamma)$ that converges to $\theta_0^* = (c_0^*, \delta_0^*, \gamma_0^*)$. However, these pseudo-true values are functions of θ_0 and f_2 . Therefore, they cannot be equal to the true value, that does not depend on f_2 . In other words, the estimator $\hat{\theta}_n = (\hat{c}_n, \hat{\delta}_n, \hat{\gamma}_n)$ converges to a stochastic limit whose distribution depends on the distribution of f_2 .

(ii) Time series asymptotics: n fixed, $T \to \infty$

This is the multivariate time series framework. The standard results for the consistency of estimator $\hat{\theta}_T$ in state space models apply [see e.g. Fuh (2006)].

(iii) Double asymptotics: $n \to \infty, T \to \infty$

Let us now consider the double asymptotics with CL(1) approach. We have:

$$L_{cc,n,T}(\theta) = \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=1}^{T} [n_{kl,t} \log p_{kl}(\theta)].$$

Since T is now varying, we need uniform a.s. convergence of the ratios $(n_{l,t}/n), l = 1, ..., K$, and $(n_{k,l,t}/n_{l,t}), k, l = 1, ..., K$, with respect to t, not only their pointwise a.s. convergence. A sufficient condition for this uniform convergence is

Assumption A.5: $n, T \to \infty$ with $T/n \to 0$.

This corresponds to the panel estimation with the cross-sectional dimension much larger than the time dimension. Under this Assumption, the uniformity in t is easily checked [see Appendix C].

By the property of the Kullback-Leibler divergence measure, we know that this limiting composite log-likelihood is maximized iff

$$p_{kl}(\theta) = p_{kl}(\theta_0), \forall k, l.$$

Then, by the identifiability of $\theta=(c,\delta,\gamma)$ (see Proposition 1), we get $\theta_0^*=\theta_0$, and the consistency follows.

4.3 Asymptotic Normality

For expository purpose, we continue the discussion of the CL(1) approach. As noted above, the composite log-likelihood is continuously differentiable. Since the estimator $\hat{\theta}_{n,T} = (\hat{c}_{n,T}, \hat{\delta}_{n,T}, \hat{\gamma}_{n,T})$ tends to the true value $\theta_0 = (c_0, \delta_0, \gamma_0)$, which is in the interior of the parameter set, the estimator will also be asymptotically in the interior of the parameter set and will satisfy the necessary first-order conditions for large T. Therefore, we have:

$$\frac{\partial L_{cc,n,T}(\hat{\theta}_{n,T})}{\partial \theta} = 0 \iff \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=1}^{T} \left[n_{kl,t} \frac{\partial log p_{kl}(\hat{\theta}_{n,T})}{\partial \theta} \right] = 0.$$

We can perform a Taylor-McLaurin expansion with respect to $\hat{\theta}_{n,T}$ in the neighborhood of θ_0 . Let us assume:

Assumption A.6: The parameter set Θ for θ is convex.

We get:

$$\sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=2}^{T} \left[n_{kl,t} \frac{\partial log \ p_{kl}(\theta_0)}{\partial \theta} \right] + \left(\sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=2}^{T} \left[n_{kl,t} \frac{\partial^2 log \ p_{kl}(\tilde{\theta}_{n,T})}{\partial \theta \ \partial \theta'} \right] (\hat{\theta}_{n,T} - \theta_0) \right) = 0, \tag{4.1}$$

where $\tilde{\theta}_{n,T}$ is an intermediate value between $\hat{\theta}_{n,T}$ and θ_0 .

By applying the same argument as for the uniform a.s. convergence of the log-likelihood function, we deduce that:

$$\frac{1}{nT} \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=2}^{T} \left[n_{kl,t} \frac{\partial^{2} log \ p_{kl}(\tilde{\theta}_{n,T})}{\partial \theta \partial \theta'} \right] \text{ will converge a.s. to } \sum_{k=1}^{K} \sum_{l=1}^{K} \left[p_{l}(\theta_{0}) p_{kl}(\theta_{0}) \frac{\partial^{2} log \ p_{kl}(\theta_{0})}{\partial \theta \partial \theta'} \right],$$

$$\frac{1}{nT} \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=2}^{K} \left[n_{k,l,t} \frac{\partial log p_{kl}(\theta_{0})}{\partial \theta} \right] \text{ will converge a.s. to } \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=2}^{T} \left[p_{l}(\theta_{0}) p_{kl}(\theta_{0}) \frac{\partial log \ p_{kl}(\theta_{0})}{\partial \theta} \right],$$
and

$$\frac{1}{n\sqrt{T}} \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{t=2}^{T} \left[n_{kl,t} \frac{\partial log \ p_{kl}(\theta_0)}{\partial \theta} \right] = \frac{1}{\sqrt{T}} \sum_{k=1}^{K} \sum_{l=1}^{K} \left[\left(\frac{1}{n} \sum_{t=2}^{T} n_{kl,t} \right) \frac{\partial log \ p_{kl}(\theta_0)}{\partial \theta} \right] \\
= \frac{1}{\sqrt{T}} \sum_{k=1}^{K} \sum_{l=1}^{K} \left\{ \left[\sum_{t=1}^{T} p_{l,t}(\theta_0) p_{kl,t}(\theta_0) \right] \frac{\partial log \ p_{kl}(\theta_0)}{\partial \theta} \right\} + o_p(1),$$

where $o_P(1)$ is a negligible term in probability. Let us assume:

Assumption A7: The matrix $J_0 = \sum_{k=1}^K \sum_{l=1}^K \left[p_l(\theta_0) p_{kl}(\theta_0) \frac{\partial^2 log \ p_{kl}(\theta_0)}{\partial \theta \partial \theta'} \right]$ is positive definite.

Then, by normalizing the expansion (4.1) by $1/(n\sqrt{T})$, we get:

$$\sqrt{T}(\hat{\theta}_{n,T} - \theta_0) = \left[-\sum_{k=1}^K \sum_{l=1}^K \left(p_l(\theta_0) \ p_{kl}(\theta_0) \ \frac{\partial^2 \log \ p_{kl}(\theta_0)}{\partial \theta \partial \theta'} \right) \right]^{-1} \\
\times \frac{1}{\sqrt{T}} \sum_{t=2}^T \left[\left(vec \left(p_{l,t}(\theta_0) p_{kl,t}(\theta_0) \right) \right)' \times \left(vec \left(\frac{\partial \log \ p_{kl}(\theta_0)}{\partial \theta} \right) \right) \right] + o_p(1), \tag{4.2}$$

where vec denotes the vectorialization with respect to the indexes k, l.

The common factor f_t is strictly stationary, geometrically mixing. Thus, the same property holds for the K^2 dimensional process $vec(p_{l,t}(\theta_0) p_{kl,t}(\theta_0))$, as well as for the $dim(\theta)$ dimensional process $[vec(p_{l,t}(\theta_0) p_{kl,t}(\theta_0))]' \left[vec\left(\frac{\partial \log p_{kl}(\theta_0)}{\partial \theta}\right)\right]$. Moreover, we have:

$$E_{0}\left[\left(vec\left(p_{l,t}(\theta_{0})p_{kl,t}(\theta_{0})\right)'vec\left(\frac{\partial \log p_{kl}(\theta_{0})}{\partial \theta}\right)\right]$$

$$=E_{0}\left[\sum_{k=1}^{K}\sum_{l=1}^{K}\left(p_{l}(\theta_{0})p_{kl}(\theta_{0})\frac{\partial \log p_{kl}(\theta_{0})}{\partial \theta}\right)\right]$$

$$=\frac{\partial}{\partial \theta}\left(E_{0}\left[\sum_{k=1}^{K}\sum_{l=1}^{K}p_{l}(\theta_{0})p_{kl}(\theta_{0})logp_{kl}(\theta)\right]\right)_{\theta=\theta_{0}}=0,$$

since we can commute the derivative $\frac{\partial}{\partial \theta}$ and the expectation (as $\frac{\partial log p_{kl}}{\partial \theta}(\theta)$ is uniformly bounded on Θ) and since θ_0 is the solution of the asymptotic optimization. Therefore, we get the asymptotic normality of:

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \left[vec\left(p_{l,t}(\theta_0) p_{kl,t}(\theta_0)\right)' vec\left(\frac{\partial log \ p_{kl}(\theta_0)}{\partial \theta}\right) \right],$$

by the geometric ergodicity of the $p'_{l,t}$ s, and, by equation (4.2), the asymptotic normality of

$$\sqrt{T}(\hat{\theta}_{n,T} - \theta_0),$$

follows.

Proposition 3 Under Assumptions A.1-A.7, the maximum conditional composite likelihood estimator $\hat{\theta}_{n,T}$ obtained by maximizing $L_{cc,n,T}(\theta)$ is consistent, converges to the true value θ_0 at speed $\frac{1}{\sqrt{T}}$, and is asymptotically normal:

$$\sqrt{T}\left(\hat{\theta}_{n,T} - \theta_0\right) \sim N\left[0, J_0^{-1}\left(\sum_{h=-\infty}^{\infty} I_{0h}\right) J_0^{-1}\right],$$

where

$$J_{0} = -\sum_{k=1}^{K} \sum_{l=1}^{K} \left[p_{l}(\theta_{0}) p_{kl}(\theta_{0}) \frac{\partial^{2} log \ p_{kl}(\theta_{0})}{\partial \theta \partial \theta'} \right],$$

$$I_{0h} = \left(vec \left(\frac{\partial \ log \ p_{kl}(\theta_{0})}{\partial \theta} \right) \right)' Cov_{0} \left[vec \left(p_{l,t}(\theta_{0}) p_{kl,t}(\theta_{0}) \right), vec(p_{l,t-h}(\theta_{0}) p_{kl,t-h}(\theta_{0})) \right] \times vec \left(\frac{\partial log \ p_{kl}(\theta_{0})}{\partial \theta} \right),$$

$$h = 1, 2, ...$$

As expected, we obtained the following:

- (i) The speed of convergence of $\hat{\theta}_{n,T}$ is $\frac{1}{\sqrt{T}}$ instead of $\frac{1}{\sqrt{nT}}$ as in the granularity approximated complete log-likelihood. This is a consequence of the crude cross-sectional aggregation of the data in the composite approach as if the observations $y_{i,t}$ were cross-sectionally independent.
- (ii) The asymptotic variance is obtained from the "sandwich" formula, as it is common in a mis-specified (pseudo) maximum likelihood approach [see Hubert (1967), White (1982)].
- (iii) The terms $(p_{l,t}p_{kl,t})$ and $(p_{l,t-h}p_{kl,t-h})$ depend on f_t , f_{t-1} and f_{t-h} , f_{t-h-1} , respectively. They are correlated due to the factor dynamics (except when $\rho = 0$, that is the case of an i.i.d. factor). Therefore, the covariances have to be taken into account even if we consider

only a small number of values of lag h. Note that the sum $\sum_{h=-\infty}^{\infty} I_{0h}$ always exists due to the geometric ergodicity of the factor process.

The above asymptotic analysis is different from the main literature on composite likelihood that usually considers either i.i.d. individuals, or finite dimensional time series [see e.g. Cox, Reid (2004), Varian, Reid, Firth (2011)].

The asymptotic variance-covariance matrix of the composite maximum likelihood estimator is consistently estimated by considering appropriate sample counterparts of components J_0 , I_{0h} . The sample counterparts of J_0 and I_{0h} are described in Appendix D.1.

5 Simulation Results

In this section, we undertake a Monte Carlo experiment to assess the finite sample properties of estimators based on the conditional composite likelihood function.

5.1 The Designs

The designs include K=8 ratings, which satisfy the order conditions in Section 3.3.1. These eight states are distinguished by aggregating 14 states of credit quality used by the bank. These credit rating categories have been aggregated into 8 categories resembling the the Standard & Poor's credit ratings scale for long-term bonds, which consists of the following eight categories: AAA, AA, A, BBB, BB, B, CCC/CC, and D. The best rating AAA means an "extremely strong" capacity of the borrower to repay its debt, while the worst rating D means that the issuer of the bond is "in default". The intermediate ratings between the two extreme cases indicate a decreasing capacity to repay which corresponds to "very strong", "strong", "adequate", "faces major future uncertainties", "faces major uncertainties", and "currently vulnerable and/or has filed for a bankruptcy protection as Chapter 11", respectively. In the following, each rating is denoted by $k=1,\ldots,8$, where a higher k indicates a lower capacity to repay debt.

5.1.1 Design of the deterministic component

Given the rating at time t-1, i.e. $y_{i,t-1} = l \in \{1, ..., 7\}$. We suppose the underlying latent continuous quantitative score $y_{i,t}^*$ can be written as:

$$y_{i,t}^* = \delta_l + \beta_l f_t + \sigma_l u_{i,t}, \ u_{i,t} \sim i.i.d.N(0,1),$$

where the rating is determined by:

$$y_{i,t} = k, k = 1, \dots, 8 \iff c_k \le y_{i,t}^* < c_{k+1}, k = 1, \dots, 8,$$

with the thresholds (c_k) described in Table 1 and the intercepts (δ_l) described in Table 2.

Table 1: Thresholds (c_k)											
\overline{k}	k 1 2 3 4 5 6 7 8 9										
c_k	c_k $-\infty$ 0 1.5 3 4.5 6 7.5 9 ∞										

Table 2: Intercepts (δ_l)									
l	l 1 2 3 4 5 6 7								
δ_l	-0.5	1	2.5	4	5.5	7	8.5		

The thresholds and intercepts are ranked in an increasing order, and their values are chosen to get higher transition probabilities on the main diagonal and decreasing probabilities when a firm transits to other states. The treatment of "absorbing state D" corresponding to l=8 is discussed later on.

5.1.2 Design of the risk components

The uncertainty on migrations is driven by specific shocks $u_{i,t}$ and common systematic shocks f_t . To see the effects of the distribution of risks between the systematic and the specific components, we consider different designs for σ_l , β_l , $l=1,\ldots,7$. In order to account for the identification properties of the composite likelihood methods $\mathrm{CL}(1)$, $\mathrm{CL}(2)$ (see Propositions 1 and 2, in Section 3.3.2) and the different definitions of the transformed parameters γ_1 , we consider three designs. In the first design, appropriate for $\mathrm{CL}(1)$ (see Proposition 1), we set $\beta_1 = \sigma_1 = \frac{1}{\sqrt{2}}$, such that $\gamma_1 = \sqrt{\sigma_1^2 + \beta_1^2} = 1$. In the two other designs, appropriate for $\mathrm{CL}(2)$ (see Proposition 2), we take into account the autocorrelation parameter ρ and fix $\beta_1 = \sigma_1 = \frac{1}{\sqrt{2-\rho^2}}$, such that $\gamma_1 = \sqrt{\sigma_1^2 + (1-\rho^2)\beta_1^2} = 1$. Next, we have also to fix values of σ_l , β_l , $l \geq 2$, compatible with the fixed values of σ_1 , β_1 .

Design 1: The idiosyncratic and systematic components have, for each l, the same impact, that is: $\sigma_l = \beta_l = \frac{(1+r)^{l-1}}{\sqrt{2}}$, with r = 0.05. Thus, there is more risk when the firm is downgraded.

Design 2: The two risky components have, for each l, the same impact, that is: $\sigma_l = \beta_l = \frac{(1+r)^{l-1}}{\sqrt{2-\rho^2}}$, with r = 0.05, taking the autocorrelation parameter into account.

Design 3: The impact of the systematic component relative to the idiosyncratic one decreases with l. This means that the idiosyncratic errors largely explain the junk bonds in non crisis

environment. To capture this feature, we consider the ratios $\frac{\beta_l}{\sigma_l} = \frac{1}{(1+r)^{l-1}}$, with r = 0.05, where $\beta_l = \frac{1}{\sqrt{2-\rho^2}}$.

There is also a persistence of the systematic factor f_t , that satisfies:

$$f_t = \rho f_{t-1} + \sqrt{1 - \rho^2} \eta_t, \eta_t \sim i.i.d.N(0, 1),$$

where the autocorrelation parameter ρ measures the persistence and $f_0 \sim N(0,1)$.

We consider three values for the autocorrelation parameter ρ , that pertains are:

- i) $\rho = 0$, that corresponds to independent migration matrices. This is the basic assumption of the Value of the Firm model introduced in Vasicek (2015).
- ii) $\rho = 0.4$, that corresponds to the autocorrelation at lag 1 of the monthly Chicago Board Options Exchange volatility index (VIX) over 60 months from January 2013 to December 2018. Indeed, the VIX index is often considered as a good proxy for systematic risk factor, since it represents the investors forecasts of future market uncertainty, as reflected in the prices of derivatives written on the S&P 500.
- iii) We also consider a value of ρ , $\rho = 0.7$. This value corresponds to the autocorrelation of monthly values of VIX over the period from January 2008 to December 2008 during the financial crisis.

The autocorrelation value depends on the time unit in our discrete time model. In practice, this time unit is one month. Therefore, a value $\rho = 0.1$ corresponds to a daily autocorrelation $\rho^{1/30} \approx 0.92$, and the commonly observed "volatility clustering" in daily returns. For $\rho = 0.7$, the associated autocorrelation in daily data is about 0.99.

5.1.3 Treatment of the absorbing state

The last state, i.e. default, is an absorbing state. Therefore, if we follow a given population of corporates, all of the corporates will default at some date, and the number of still alive corporates (the so-called Population-at-Risk (PaR)) will diminish. Theoretically, the process of observed ratings is asymptotically stationary with a stationary distribution equal to a point mass on default. Hence, the regulatory conditions for the convergence and asymptotic normality of the maximum composite likelihood estimators are not satisfied. This difficulty can be (partly) solved in two alternative ways.

i) There is currently a change of regulation known as the resolution step. The idea is to assist the very risky firms before they enter into default. Loosely speaking, in the risky grade 7, the supervisory authority can partly take control of the firm and monitor its restructuring and its debt renegotiation to avoid default, when default is due to transitory difficulties. This reduces the probability of default, while avoiding

this probability being equal to zero. We do not incorporate the resolution step in our modelling, as that would require a modelling of the supervisory authority behaviour.

ii) The second approach assumes that newly created corporates offset the corporates entering into default, thus ensuring a PaR of constant size. This corresponds to the model with equal birth and death rates used in epidemiological studies (see e.g. Harko, Lobo, Mak (2016)). As at the time of new firms arrival their rating are high, we replace the last row of the migration matrix at the individual level,

corresponding to a standard absorbing state, by the row of assignment of new entries at the population level,

Thus we have to distinguish individual migration matrices P_t , from the population migration that can be or not adjusted by taking into account the newly created firms. When the newly created firms are taken into account, the migration matrix will be indexed as P_t^a .

5.1.4 Individual trajectories

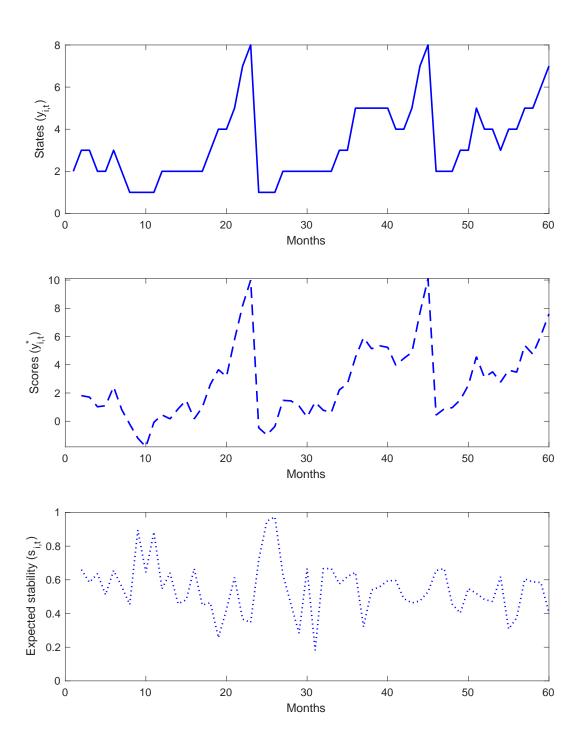
Let us for instance consider the design 3 with $\rho = 0.4$. For each individual i, we can compare different time series, such as the series of the underlying scores, of the ratings as well as the series providing the expected stability in the current score. The series are $y_{i,t}^*$, $y_{i,t}$ and $s_{i,t}$, where

$$s_{i,t} = \Phi\left(\frac{c_{l+1} - \beta_l f_t + \delta}{\sigma_l}\right) - \Phi\left(\frac{c_l - \beta_l f_t + \delta}{\sigma_l}\right), \text{ with } l = y_{i,t-1}.$$

We can also consider the similar series, with $f_t = 0$, for all t, that is without systematic factor, in order to provide insights on the effect of f_t . Such trajectories are provided in Figure 2 with an initial factor value set at $f_1 = 0$, and an initial rating set at $y_{i,0} = 2$ corresponding to AA. The trajectories correspond to three different corporate bonds. At time 0, a bond is issued with a rating 2 (AA). It has some downgrading after month 10 up to default at month 21. At this time a new bond is issued to balance the defaulted bond at a rating 1 (AAA), with a systematic downgrading up to default at month 44. Then a new bond is issued at time 45 and so on.

In such an equal birth-death rates environment, each trajectory corresponds to a stochastic number of firms, not to a single firm. This stochastic number is equal to the number of times default is reached plus one. The advantage of this practice is in ensuring the stationarity of the process, and also getting the rating histories of equal length T.

Figure 2: Monthly Individual Trajectories, Design 3, $\rho=0.4$



5.1.5 Migration Matrices

In this section we present the matrices for design 3, $\rho = 0.4$. Recall that the time unit is one month and the horizon one and two refer to the periods of one and two months, respectively. Matrix P^a is given in Table 3.

P^a	k=1	k=2	k=3	k=4	k=5	k = 6	k = 7	k = 8
l=1	68.42	28.82	2.72	0.04	0.00	0.00	0.00	0.00
l=2	17.48	50.53	28.93	3.01	0.05	0.00	0.00	0.00
l = 3	1.14	16.97	49.46	29.01	3.35	0.07	0.00	0.00
l=4	0.02	1.31	17.43	48.36	29.07	3.71	0.10	0.00
l = 5	0.00	0.03	1.53	17.88	47.23	29.09	4.11	0.13
l = 6	0.00	0.00	0.04	1.78	18.32	46.07	29.07	4.72
l = 7	0.00	0.00	0.00	0.06	2.07	18.73	44.89	34.25
l = 8	50.00	30.00	20.00	0.00	0.00	0.00	0.00	0.00

Table 3: Migration Matrices P^a , at Horizon 1 in %

We observe a common feature of a migration matrix, which is the main diagonal with large values and two adjacent diagonals with larger values for downgrade than the upgrade. Moreover, there are significant probabilities of default from grades 6 and 7, corresponding to "junk bonds". Note also the high probability of default for rating 1 (AAA). Then we can look for the nondegenerate stationary distribution π^a solution of:

$$(\pi^a)' = (\pi^a)' P^a. (5.1)$$

As noted earlier, due to the absorbing state, and without equal birth-death rates, each corporate bond will default and the asymptotic stationary distribution at the individual level would be a point mass at 8 (D). The interpretation of stationary distribution π^a is different and at the population level. It provides the long run rating structure of the population of corporate bonds under rebalancing. This long run structure is given in Table 4. In practice, when analyzing the ratings, the stationary distribution provides information on how the ratings agencies are discretizing their scores to define the ratings. In our experimental design, they are discretized to get rather similar proportions of bonds rated $1, \ldots, 7$.

Table 4: Stationary Distribution

	k=1	k=2	k = 3	k=4	k = 5	k = 6	k = 7	k = 8
Probabilities in %	14.51	16.66	17.47	16.09	14.15	11.19	6.99	2.94

Let us now consider the migration matrix at horizon 2. Table 5 shows the matrices $P^a(2)$ and $(P^a)^2$. $(P^a)^2$ is computed under closed form and $P^a(2)$ computed by Monte-Carlo integration

with S=50,000 replications (see Lemma 3). The changes in $P^a(2)$ and $(P^a)^2$ comparatively to P^a are due to variation in the aggregate effect of both specific and systematic shocks. Both matrices have now non zero elements on the diagonals distant by 2 from the main diagonal by time aggregation. The matrices $P^a(2)$ and $(P^a)^2$ are not equal. Their difference captures the effect of the systematic uncertainty. Next we repeat the same exercise, when there is no systematic effect, that is if $f_t=0$, for all t, or equivalently $\beta_l=0$, for all l. The results are presented in Table 6. As expected $P^a(2)$ and $(P^a)^2$ coincide.

Table 5: Migration Matrices, at Horizon 2 in %

$P^a(2)$	k=1	k=2	k=3	k=4	k=5	k = 6	k = 7	k = 8
$\overline{l=1}$	51.89	34.75	11.54	1.70	0.12	0.00	0.00	0.00
l=2	21.12	35.51	29.93	11.39	1.90	0.15	0.00	0.00
l=3	4.31	17.68	34.51	29.49	11.69	2.12	0.19	0.01
l=4	0.45	4.27	17.88	33.75	29.05	11.99	2.36	0.25
l = 5	0.09	0.56	4.64	18.06	32.97	28.58	12.26	2.84
l = 6	2.36	1.45	1.57	4.99	18.21	32.07	27.20	12.15
l = 7	17.13	10.28	6.90	0.76	5.35	17.64	25.68	16.26
l = 8	39.68	32.96	19.93	6.73	0.69	0.01	0.00	0.00
P^a	k = 1	k=2	k = 3	k=4	k = 5	k = 6	k = 7	k = 8
$\frac{(P^a)^2}{l=1}$	k = 1 52.90	k = 2 31.85	k = 3 12.59	k = 4 2.40	k = 5 0.25	k = 6 0.01	k = 7 0.00	k = 8 0.00
$\overline{l=1}$	52.90	31.85	12.59	2.40	0.25	0.01	0.00	0.00
$ \begin{array}{c} l = 1 \\ l = 2 \end{array} $	52.90 22.83	31.85 33.32	12.59 28.37	2.40 12.56	0.25 2.61	0.01 0.29	0.00 0.02	0.00
$ \begin{array}{c} l = 1 \\ l = 2 \\ l = 3 \end{array} $	52.90 22.83 5.61	31.85 33.32 17.88	12.59 28.37 32.51	2.40 12.56 28.06	0.25 2.61 12.74	0.01 0.29 2.83	0.00 0.02 0.35	0.00 0.00 0.02
$ \begin{array}{c} \overline{l} = 1 \\ l = 2 \\ l = 3 \\ l = 4 \end{array} $	52.90 22.83 5.61 0.76	31.85 33.32 17.88 5.23	12.59 28.37 32.51 18.03	2.40 12.56 28.06 31.82	0.25 2.61 12.74 27.72	0.01 0.29 2.83 12.92	0.00 0.02 0.35 3.08	0.00 0.00 0.02 0.44
	52.90 22.83 5.61 0.76 0.13	31.85 33.32 17.88 5.23 0.86	12.59 28.37 32.51 18.03 5.56	2.40 12.56 28.06 31.82 18.16	0.25 2.61 12.74 27.72 31.13	0.01 0.29 2.83 12.92 27.33	0.00 0.02 0.35 3.08 13.09	0.00 0.00 0.02 0.44 3.74

Table 6: Migration Matrix at Horizon 2 without Systematic Risk

$P^a(2) = (P^a)^2$	k=1	k=2	k=3	k=4	k=5	k = 6	k = 7	k = 8
l=1	58.84	34.25	6.70	0.21	0.00	0.00	0.00	0.00
l=2	13.71	46.46	32.23	7.28	0.31	0.01	0.00	0.00
l=3	1.21	13.75	44.50	32.19	7.90	0.44	0.01	0.00
l=4	0.03	1.50	14.65	42.66	32.00	8.53	0.61	0.02
l = 5	0.00	0.05	1.86	15.46	40.92	31.68	9.17	0.86
l = 6	0.84	0.50	0.42	2.27	16.19	39.27	30.97	9.54
l = 7	15.32	9.19	6.13	0.14	2.73	16.61	33.15	16.73
l = 8	40.53	33.72	20.22	5.39	0.14	0.00	0.00	0.00

5.2 Finite Sample Properties of the MCL Estimation

To give some insights into the accuracy of the MCL estimation, in terms of the number of months T and the factor autocorrelation parameter ρ , we conduct some Monte-Carlo experiments. The estimation is performed with N=1,000 firms, including the adjustment for the newly created firms, and for the different designs described above. The number of observation periods is fixed to T=60 (5 years), T=120 (10 years), T=240 (20 years). For each experiment, we perform S=1,000 simulations of individual trajectories, with initial ratings $y_{i,0}$ drawn from the adjusted stationary distribution π^a , conditional to nondefault ratings.

5.2.1 Parameters of Interest

The stochastic migration model depends on a large number of parameters that are 19 identifiable parameters $(c_3 \ldots, c_8, \delta_1 \ldots, \delta_7, \gamma_2 \ldots \gamma_7)$ for the CL(1), and 27 identifiable parameters $(c_3 \ldots, c_8, \delta_1 \ldots, \delta_7, \beta_2 \ldots, \beta_7, \sigma_1 \ldots, \sigma_7)$ for the CL(2), taking into account the two identification conditions $c_2 = 0, \gamma_1 = 1$. We discuss the estimation of these parameters, and use their estimates to illustrate the prediction of the following downgrade probabilities and probabilities of default.

- i) the downgrade probabilities at horizon 1 and 2 of a firm currently rated A (l=3): DP(1|A), DP(2|A)
- ii) the term structure of the probability of default at different horizons h for a firm currently rated A. The horizons are fixed to 1 month, 1 year, 2 years, 3 years, and denoted PD(1|A), PD(12|A), PD(24|A), PD(36|A)

In the next sections, we discuss the CL(1) estimation results based on design 1, and the CL(2) estimation results based on designs 2 and 3.

5.2.2 The CL(1) Estimation Results

Table 7 in Appendix D.2 shows the accuracy of the CL(1) estimates for design 1 when the autocorrelation parameters $\rho = 0$. In the table, we report for each identifiable parameter, its true value, the mean absolute bias of the CL(1) estimates, and the associated standard errors in terms of the number of months T. The absolute mean bias is computed by averaging the absolute value of the bias. To find the standard errors, we compute the estimator $\hat{\Sigma}_{\theta} = \widehat{Var}\left(\sqrt{T}\left(\hat{\theta}_{nT} - \theta\right)\right)$ of the asymptotic covariance matrix $\Sigma_{\theta} = J_0^{-1}\left(\sum_{h=-\infty}^{\infty} I_{0h}\right)J_0^{-1}$ in Proposition 3, which is the heteroskedasticity and autocorrelation consistent (HAC) estimator. The HAC estimator is obtained using a quadratic spectral kernel and a bandwidth,

which we set to be $4(T/100)^{2/9}$ following Newey, West (1994). The standard error is then obtained from the diagonal elements of the average of the estimated variance over all simulations after dividing by T.

Table 7 shows that the CL(1) estimates have a small estimated bias and standard errors relative to their true value. Both decrease as the number of months increases. For instance, when $\rho = 0$, the largest absolute bias is 1.3, when estimating the threshold parameter c_8 , and T = 60, which remains small relative to its true value $c_8 = 9$. This bias decreases to 0.66 when T = 240. Meanwhile, the standard error decreases from 1.20 to 0.60. These results are consistent with the asymptotic results on the \sqrt{T} -consistency of the CL(1) estimates in Proposition 3. Also, we notice that, in general, there is a small effect of ρ on the consistent estimation of the different parameters. Tables 8 and 9 show that when the autocorrelation parameter increases to $\rho = 0.4$ and $\rho = 0.7$, the bias and standard errors remain small. When ρ increases, there is more persistence in f_t and then less information on the shocks on f. This impacts the accuracy that diminishes when ρ increases, while being still compatible with the consistency of the estimation.

Next, we illustrate the finite sample performance of the inferences based on the standard asymptotically valid t-statistics by plotting the empirical probability distribution function (PDF) of each computed t statistic when testing that each of the parameters equals their true value. See Figures 3-11 in Appendix D.2. In Figures 3-5, we plot the empirical probability distribution by depicting the histogram of the computed test statistics over the S simulations for the threshold parameters (c_3, \ldots, c_8) , when the time dimensions change for the different values of ρ . Figures 6 – 8 show the empirical probability distributions for the intercepts $(\delta_1, \ldots, \delta_7)$, while Figures 9 – 11 present the distributions for the unconditional variances $(\gamma_2, \ldots, \gamma_7)$. In each figure, the x-axis represents the computed t statistic over the simulations, while the y-axis contains their frequencies. A common feature in the figures is that, when T varies, the distribution of t statistics is centered on zero. In most cases, the simulated t-statistics belong to the intervals [-1.96, 1.96], as would be expected for asymptotically normally distributed t statistics, when the level of the test is 5%. However, we observe on Figures 6-8, that when the autocorrelation parameter becomes higher ($\rho = 0.4$ and $\rho = 0.7$), for some of the δ_l parameters, the realized t statistics are shifted from the zero mean due to the smaller variance, suggesting an under-estimation of the variance of δ_l . The shift is negative for the negative values of parameters δ_l and is positive for the positive values of the parameters.

5.2.3 The CL(2) Estimation Results

We provide in Table 10 (see Appendix D.3) the mean absolute bias and the standard errors of the estimates of $c_k, k = 3, ..., 8$, $\delta_l, l = 1, ..., 7$, $\beta_l, l = 2, ..., 7$, $\sigma_l, l = 1, ..., 7$ and ρ , for design 2 when $\rho = 0$. The estimate for β_1 is deduced from $\beta_1 = \sqrt{\frac{1-\sigma^2}{1-\rho^2}}$ satisfying the condition $\gamma_1 = 1$ and taking into account the sign restriction $\beta_1 > 0$. The table reports

for each parameter its true value, the mean absolute bias and the standard error of the CL(2) estimation for each value of T. The main conclusion is that for each value of the time period T, the estimates have, in general, small estimated biases and standard errors for the thresholds, the intercepts, and the autocorrelation parameter. The biases and the standard errors decrease as the number of periods increases, confirming the asymptotic results in Proposition 3. For instance, when estimating c_6 , the bias decreases from 0.16 for T=60 to 0.091 for 240, and are small relative to the true value of the parameter, which is 6. The standard error decreases from 0.72 to 0.36. We observe similar results for the autocorrelation parameter, the estimated factor sensitivities, and the volatilities.

Note however that the accuracy improves slowly with T for the parameters β_l , σ_l , especially for the junk bonds rating. This is related with the nonidentification of β_l , σ_l in CL(1). Even with CL(2) the two parameters are still weakly identifiable. This could be improved by applying a two-step procedure, in which CL(1) is applied to the parameters identifiable by CL(1). Next, these estimates are plugged in the CL(2) criterion, that is next optimized w.r.t. to the β_l , σ_l . This two-step approach has not been illustrated in our simulation to avoid a large number of nested simulations.

We also observe that the estimated autocorrelation parameter is very accurate when $\rho=0$. Nevertheless, the biases and standard errors are relatively slightly higher for these parameters when the persistence in the systematic factor increases, but reduce as T rises (see Tables 11 and 12). Tables 14, 15, and 16 are about the mean absolute biases and the standard errors of the CL(2) estimators when the impact of the systematic component relative to the idiosyncratic one decreases with l in design 3. The findings remain the same as for design 2.

Some of the parameters introduced in the model have important interpretations. This is typically the case of the threshold parameters c_l . The other parameters do not necessarily have such direct practical interpretation. In fact, the real parameters of interest are generally nonlinear function of β_l , σ_l , for which the Monte-Carlo analysis is not provided in the previous tables. For instance, we might be interested in the relative weight of the idiosyncratic and systematic shocks, that is the ratio β_l/σ_l , whose accuracy cannot be deduced from the tables which provides no information on the dependence between $\hat{\beta}_l$ and $\hat{\sigma}_l$. Other parameters of interest are the downgrade probabilities and the probability of default. They will be analyzed in the next paragraph.

Tables 13 and 17 present the downgrade probabilities and the probabilities of default at different horizons for design 2 and design 3, respectively. The downgrade probabilities and probabilities of default are computed from the average over S Monte-Carlo approximations of the probabilities of transition at horizon h, $P(h) = E(P_t P_{t-1} \dots P_{t-(h-1)})$, after plugging in the estimated parameters. Let us first discuss the results for design 2 when $\rho = 0$ in the first panel of Table 13. We noted that the downgrade probability and the probability of default increase as the horizon increases, and this feature is reproduced after the estimation. At horizon 1, when T = 60, the downgrade probability is 33.29%, while its true value is

32.52%. At horizon 2, when T=60, the estimated probability increases to 44.67%, which is also close to the true value, 43.35. At horizon 1, all the probabilities of default are very close to zero for all values of T. This is consistent with the fact that the probability for a firm with a strong capacity to repay its debt (a firm rated A) defaults is negligible within one month. The estimated probabilities are closer to their true values when the autocorrelation parameter is small. See also the second panel and the third panel of Table 13, illustrating design 2 and design 3 when $\rho=0.4$ and $\rho=0.7$, respectively. The findings for design 3, presented in Table 17, are very similar.

6 Conclusion

This paper proposes a MCL estimation method for the stochastic factor ordered Probit model, as an alternative to other approximation-based estimation methods validated by the banking supervisory authority. The advantage of this method is that it can be easily adjusted to a varying numbers of credit quality categories for internal credit rating analysis. In addition, it allows for introducing one or possibly multiple latent factors to capture the common dynamics of credit quality of the borrowers or of a leading indicator on their market. Because, the conditional migration matrices are functions of the parameters θ as well as of the common factor values $(f_t), t = 1, \dots, T$, the log-likelihood function contains a multivariate integral of order T. We propose three MCL estimators of different complexity: conditional composite log-likelihood function at lag 1, the conditional composite log-likelihood at lag 2, and the conditional composite likelihood up to lag 2. The paper discusses the identifiability of the model parameters and establishes the asymptotic properties of the maximum conditional composite likelihood estimators. It derives the conditions for the uniform convergence with respect to the identifiable parameters, and for the asymptotic normality of the proposed estimators. We illustrate the finite sample properties of the conditional composite loglikelihood at lag 1 and the conditional composite log-likelihood at lag 2 through Monte-Carlo experiments.

Appendix A: The Expected Transition Matrices

A.1. Expected Matrix P (Lemma 1)

We have:

$$y_{i,t}^* = \beta_l f_t + \delta_l + \sigma_l u_{i,t}$$
, if $y_{i,t-1} = l$,

where $u_{i,t} \sim N(0,1)$ and $f_t \sim N(0,1)$ are independent. Then, if $y_{i,t-1} = l$, $y_{i,t}^* \sim N(\delta_l, \sigma_l^2 + l)$

 β_l^2). It follows that:

$$P[y_{i,t} = k | y_{i,t-1} = l] = P[c_k < y_{i,t}^* < c_{k+1} | y_{i,t-1} = l],$$

and

$$p_{kl}(\theta) = \Phi\left(\frac{c_{k+1} - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2}}\right) - \Phi\left(\frac{c_k - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2}}\right).$$

A.2. Matrix P(2) (Lemma 2)

We have:

$$P(2) = E[P(f_t; \theta) \ P(f_{t-1}; \theta)] = E[P(\rho f_{t-1} + \sqrt{1 - \rho^2} \ \eta_t; \theta)] \ P(f_{t-1}; \theta)].$$

Since f_{t-1} and η_t are independent, $\eta_t \sim N(0,1)$ and $f_{t-1} \sim N(0,1)$, we get:

$$P(2) = E_{f_{t-1}} E_{\eta_t} \left[P(\rho f_{t-1} + \sqrt{1 - \rho^2} \eta_t; \theta) P(f_{t-1}; \theta) | f_{t-1} \right],$$

$$= E_{f_{t-1}} \left[E_{\eta_t} \left[P(\rho f_{t-1} + \sqrt{1 - \rho^2} \eta_t; \theta) | f_{t-1} \right] P(f_{t-1}; \theta) \right],$$

$$= E_{f_{t-1}} \left[A B \right],$$

where the components of matrix A are given by:

$$a_{kl}(f_{t-1}; \theta, \rho) = \mathbb{P}\left[c_k < y_{i,t}^* < c_{k+1} | y_{i,t-1} = l, f_{t-1}\right]$$

$$= \mathbb{P}\left[c_k < \delta_l + \beta_l \rho f_{t-1} + \beta_l \sqrt{1 - \rho^2} \eta_t + \sigma_l u_{i,t} < c_{k+1} | y_{1,t-1}, f_{t-1}\right],$$

$$= \Phi\left(\frac{c_{k+1} - \delta_l - \beta_l \rho f_{t-1}}{\sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)}}\right) - \Phi\left(\frac{c_k - \delta_l - \beta_l \rho f_{t-1}}{\sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)}}\right), k, l, = 1, ...K,$$

by the independence between $(\eta_t, u_{i,t})$ and $(y_{i,t-1}, f_{t-1})$. By (2.4) the elements of matrix B are:

$$p_{kl}(f_{t-1};\theta) = \Phi\left(\frac{c_{k+1} - \delta_l - \beta_l f_{t-1}}{\sigma_l}\right) - \Phi\left(\frac{c_k - \delta_l - \beta_l f_{t-1}}{\sigma_l}\right), k, l = 1, ..., K.$$

Therefore, by integrating out f_{t-1} , we get:

$$p_{kl}(2;\theta,\rho) = \int \sum_{j=1}^{K} [a_{k,j}(f;\theta,\rho) \ p_{j,l}(f;\theta)] \phi(f) df$$

$$= \int \sum_{j=1}^{K} \left[\Phi\left(\frac{c_{k+1} - \delta_j - \beta_j \rho f}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}}\right) - \Phi\left(\frac{c_k - \delta_j - \beta_j \rho f}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}}\right) \right]$$

$$\times \left[\Phi\left(\frac{c_{j+1} - \delta_l - \beta_l f}{\sigma_l}\right) - \Phi\left(\frac{c_j - \delta_l - \beta_l f}{\sigma_l}\right) \right] \phi(f) df.$$

Appendix B: Proof of Propositions 1 and 2

B.1. Proof of Proposition 1

From Lemma 1, and the definitions $c_1 = -\infty$, $c_{K+1} = \infty$, we know that the identifying functions of parameters are:

$$\frac{c_k - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2}} \quad \forall \ k = 1, ..., K - 1, \ l = 1, ..., K.$$
 (b.1)

Therefore parameter ρ is not identifiable. Moreover, parameters σ_l^2 cannot be distinguished from β_l^2 . Let us denote their sum by γ_l^2 , where

$$\gamma_l = \sqrt{\sigma_l^2 + \beta_l^2}.$$

There are K(K-1) identifying functions (a.1), that we would like to use to identify the (K-1) values of c_k , the K values of δ_l and the K values of γ_l , i.e. 3K-1 unknowns. We follow Gagliardini, Gouriéroux (2015) and add the identifying constraints:

$$c_2 = 0, \quad \gamma_1 = 1.$$

Next, we proceed as follows:

- (a) From (b.1) written for k=2, we identify $\frac{\delta_l}{\gamma_l}$. Given that $\gamma_1=1$, we get δ_1 identified.
- (b) For l = 1, we have $\gamma_1 = 1$, hence we identify $c_k \delta_1$, given (b.1). Therefore, all thresholds c_k , k = 2, ..., K are identified.

(c) Then the identifying functions can also be written as:

$$\frac{c_k - \delta_l}{\gamma_l} = \frac{c_k}{\gamma_l} - \frac{\delta_l}{\gamma_l}, k = 2, \dots, K, l = 1, \dots, K.$$

Therefore, from the identification of the ratios δ_l/γ_l result in (a), we identify all ratios c_k/γ_l . Then from the identification of the c_k 's (b), we identify γ_l , l=1,...,K. Now, the c_k , γ_l are identified, and from (c), we identify δ_l , l=1,...,K.

B.2. Proof of Proposition 2

We have the following identifying functions of parameters:

(1)
$$\frac{c_k - \delta_l}{\sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)}}, k = 2, \dots, K, l = 1, \dots, K$$

(2)
$$\frac{\epsilon \beta_l \rho}{\sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)}}, l = 1, \dots, K$$

(3)
$$\frac{c_k - \delta_l}{\sigma_l}, k = 2, \dots, K, l = 1, \dots, K$$

$$(4) \ \frac{\epsilon \beta_l}{\sigma_l}, l = 1, \dots, K.$$

Let us define:

$$\gamma_l = \sqrt{\sigma_l^2 + \beta_l^2 (1 - \rho^2)},$$

and use the identifying constraints:

$$\gamma_1 = 1, \quad c_2 = 0.$$

Then we proceed as follows:

- (a) For k = 2, given $c_2 = 0$ and equation (1), we identify $\frac{\delta_l}{\gamma_l}$.
- (b) For k=2, given $c_2=0$, and equation (3), we identify $\frac{\delta_l}{\sigma_l}$, $l=1,\ldots,K$.
- (c) Given that $\gamma_1 = 1$, it follows from (a) that parameter δ_1 is identified.
- (d) Then, it follows from (b) that parameter σ_1 is identified.

(e) For l = 1 and equation (1), we identify

$$\frac{c_k - \delta_1}{\gamma_1} = c_k - \delta_1.$$

Hence, from (c), it follows that c_k , k = 1, ..., K - 1 are identified.

(f) From equation (1), the quantities

$$\frac{c_k}{\gamma_l} - \frac{\delta_l}{\gamma_l}$$

are identified since $\gamma_1 = 1$.

Then, by (a), the ratios $\frac{c_k}{\gamma_l}$ are identified.

- (g) From (f) and (e), parameters γ_l , l = 1, ..., K are identified.
- (h) From (a) and (g), parameters δ_l , l = 1, ..., K are identified.
- (i) From (b) and (h), parameters σ_l , l = 1, ..., K are identified.
- (j) From equation (4) and result (i), parameters $\epsilon \beta_l$, l = 1, ..., K are identified.
- (k) From (2), we get the ratios $\frac{\epsilon \beta_l \rho}{\gamma_l}$ and given (g) we identify $\epsilon \beta_l \rho$, l = 1, ..., K
- (l) Finally, from (j) and (k), we identify parameter ρ .

Appendix C: Proof of Uniform a.s. Convergence

Let us introduce a more precise notation: $\hat{p}_{k,l,t}(n,T) = (n_{k,l,t}/n_{l,t})$, where the indexes n,T are introduced to explicit the dependence in the number of individuals n and the number of dates T. Indeed, by the Hajek, Renyi inequality [Hajek, Renyi (1955), Csorgo (1967, ineq (2.8))], we get:

$$P[Max_{m\geq n}|\hat{p}_{k,l,t}(m,T) - p_{kl}(f_t,\theta_0)| > \epsilon] < \frac{1}{\epsilon^2} \sum_{m=n}^{\infty} \frac{1}{m^2} p_{kl}(f_t,\theta_0), \ \forall \epsilon > 0,$$

$$\iff P[Max_{m\geq n}|\hat{p}_{k,l,t}(m,T) - p_{kl}(f_t,\theta_0)| > \epsilon] < \frac{c}{\epsilon^2 n} p_{kl}(f_t,\theta_0), \ \forall \epsilon > 0,$$

where c is a constant. Then, it follows that:

$$P[Max_{t \le T} Max_{m \ge n} | \hat{p}_{k,l,t}(m,T) - p_{kl}(f_t,\theta_0)| > \epsilon] < \frac{c}{\epsilon^2 n} \sum_{t=2}^{T} p_{kl}(f_t,\theta_0).$$

For n, T large, the upper bound: $\frac{c}{\epsilon^2} \frac{T}{n} \frac{1}{T} \sum_{t=2}^{T} p_{kl}(f_t, \theta)$ is equivalent to $\frac{c}{\epsilon^2} \frac{T}{n} E_0[p_{kl}(f_t, \theta_0)]$, by the geometric ergodicity of factor (f_t) . Then by Assumption A5, we infer:

$$\lim_{T \to \infty} P[Max_{t \le T} Max_{m \ge n} | \hat{p}_{k,l,t}(m,T) - p_{kl}(f_t, \theta_0) | > \epsilon] = 0,$$

and the required uniformity.

Therefore, after the normalization, the a.s. limit of the normalized composite log-likelihood is:

$$\begin{split} & \lim_{n,T \to \infty} \text{a.s. } \frac{1}{nT} \sum_{k=1}^K \sum_{l=1}^K \sum_{l=1}^T \left[n_{kl,t} \log p_{kl}(\theta) \right] \\ &= \lim_{n,T \to \infty} \text{a.s. } \frac{1}{nT} \sum_{k=1}^K \sum_{l=1}^K \sum_{l=1}^K \sum_{l=1}^T \left[n_{l,t} \, \hat{p}_{kl,t} \log p_{kl}(\theta) \right] \\ &= \lim_{T \to \infty} \text{a.s. } \frac{1}{T} \sum_{l=1}^K \sum_{k=1}^K \sum_{l=1}^K \sum_{l=1}^T \left[\lim_{n \to \infty} \left(\frac{n_{l,t}}{n} \, \hat{p}_{k,l,t}(n,T) \right) \, \log p_{kl}(\theta) \right] \\ &= \lim_{T \to \infty} \text{a.s. } \frac{1}{T} \sum_{l=1}^K \sum_{k=1}^K \sum_{l=1}^K \sum_{l=1}^T \left[\mathbb{P}[c_k < y_{i,t}^* < c_{k+1}, c_l < y_{i,t-1}^* < c_{l+1} | f_t, f_{t-1}] \, \log p_{kl}(\theta) \right] \\ &(\text{by the uniform a.s. convergence}) \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[\lim_{T \to \infty} \text{a.s. } \left[\frac{1}{T} \sum_{t=2}^T \mathbb{P}[c_k < y_{i,t}^* < c_{k+1}, c_l < y_{i,t-1}^* < c_{l+1} | f_t, f_{t-1}] \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[\mathbb{P}[c_k < y_{i,t}^* < c_{k+1}, c_l < y_{i,t-1}^* < c_{l+1}] \, \log p_{kl}(\theta) \right] \text{ (since } f_t \text{ is geometrically ergodic)} \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_l(\theta_0) \, p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[p_{l0} \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta) \right] \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \left[\sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta_0) \right] \\ &= \sum_{l=1}^K \sum_{k=1}^K \sum_{k=1}^K p_{kl}(\theta_0) \, \log p_{kl}(\theta_0) \right]$$

More precisely, $L_{cc,n,T}(\theta)$ converge a.s. uniformly to

$$L_{cc,\infty}(c,\delta,\gamma) = \sum_{l=1}^{K} \left[p_{l0} \left[\sum_{k=1}^{K} p_{kl}(\theta_0) \log p_{kl}(\theta) \right] \right].$$

Appendix D: Simulation Details and Additional Results

This section provides more details for the implementation of the simulation experiments and additional results.

D.1. Simulation Details

To compute the estimated asymptotic variances, we use the estimator $\hat{\Sigma}_{\theta} = \widehat{Var} \left(\sqrt{T} \left(\hat{\theta}_{nT} - \theta \right) \right)$ of the asymptotic covariance matrix $\Sigma_{\theta} = J_0^{-1} \left(\sum_{h=-\infty}^{\infty} I_{0h} \right) J_0^{-1}$ in Proposition 3, which is the heteroskedasticity and autocorrelation consistent (HAC) estimator

$$\hat{\Sigma}_{\theta} = \hat{J}_{n,T}^{-1} \left(\hat{I}_{0,n,T} + \sum_{h=1}^{T-1} k \left(\frac{h}{B_T} \right) \left(\hat{I}_{h,n,T} + \hat{I}'_{h,n,T} \right) \right) \hat{J}_{n,T}^{-1},$$

where

$$\hat{J}_{n,T} = -\frac{1}{T} \sum_{t=2}^{T} \left(\sum_{k=1}^{K} \sum_{l=1}^{K} \frac{n_{kl,t}}{n} \frac{\partial \log \left(p_{kl} \left(\hat{\theta}_{nT} \right) \right)}{\partial \theta \partial \theta'} \right),$$

$$\hat{I}_{h,n,T} = \frac{1}{T} \sum_{t=2}^{T-h} \left(\sum_{k=1}^{K} \sum_{l=1}^{K} \left[\frac{n_{kl,t}}{n} - \frac{1}{T} \sum_{t=2}^{T} \frac{n_{kl,t}}{n} \right] \frac{\partial \log \left(p_{kl} \left(\hat{\theta}_{nT} \right) \right)}{\partial \theta} \right) \times \left(\sum_{k=1}^{K} \sum_{l=1}^{K} \left[\frac{n_{kl,t+h}}{n} - \frac{1}{T} \sum_{t=2}^{T} \frac{n_{kl,t}}{n} \right] \frac{\partial \log \left(p_{kl} \left(\hat{\theta}_{nT} \right) \right)}{\partial \theta} \right)',$$

 $k(\cdot)$ is a kernel function, and B_T is the bandwidth. The asymptotic inference is conducted using a quadratic spectral kernel with the bandwidth, which we set to $4(T/100)^{2/9}$, as suggested by Newey, West (1994). The kernel $k(\cdot)$ is a decreasing function, which accounts for the decaying dependence between the observations at t and t + h when h increases.

All derivatives were obtained using the numerical gradient function in Matlab. The second-order partial derivative in the definition of the variance estimator for the CL(2) is obtained using an outer-product argument. In particular, we note that:

$$J_{0} = -\sum_{k=1}^{K} \sum_{l=1}^{K} \left(p_{l}(\theta_{0}) p_{kl}(\theta_{0}) \left(\frac{\partial^{2} p_{kl}(\theta_{0})}{\partial \theta \partial \theta'} \frac{1}{p_{kl}(\theta_{0})} - \frac{\partial p_{kl}(\theta_{0})}{\partial \theta} \frac{\partial p_{kl}(\theta_{0})}{\partial \theta'} \frac{1}{(p_{kl}(\theta_{0})^{2})} \right) \right)$$

which is equivalent to

$$J_0 = -\sum_{l=1}^K \left(p_l(\theta_0) \frac{\partial^2 \sum_{k=1}^K p_{kl}(\theta_0)}{\partial \theta \partial \theta'} \right) + \sum_{k=1}^K \sum_{l=1}^K \left(p_l(\theta_0) p_{kl}(\theta_0) \frac{\partial p_{kl}(\theta_0)}{\partial \theta} \frac{\partial p_{kl}(\theta_0)}{\partial \theta'} \frac{1}{(p_{kl}(\theta_0)^2)} \right).$$

Since $\sum_{k=1}^{K} p_{kl}(\theta_0) = 1$, for any l,

$$J_{0} = \sum_{k=1}^{K} \sum_{l=1}^{K} \left(p_{l}(\theta_{0}) p_{kl}(\theta_{0}) \frac{\partial p_{kl}(\theta_{0})}{\partial \theta} \frac{\partial p_{kl}(\theta_{0})}{\partial \theta'} \frac{1}{(p_{kl}(\theta_{0})^{2})} \right)$$
$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \left(p_{l}(\theta_{0}) p_{kl}(\theta_{0}) \frac{\partial \log p_{kl}(\theta_{0})}{\partial \theta} \frac{\partial \log p_{kl}(\theta_{0})}{\partial \theta'} \right).$$

Hence, we estimate J_0 using

$$\hat{J}_{n,T} = \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\frac{n_{kl}}{n} \frac{\partial \log p_{kl}(\hat{\theta}_{nT})}{\partial \theta} \frac{\partial \log p_{kl}(\hat{\theta}_{nT})}{\partial \theta'} \right).$$

For the CL(2) estimators of variances, p_{kl} , n_{kl} and $n_{kl,t}$ are replaced with their lag 2 analog, respectively. In the different estimations, we replace β_1 by $\sqrt{\frac{1-\sigma^2}{1-\rho^2}}$ to comply with the identification condition $\gamma_1 = 1$ and the sign restriction on β_1 . As explained before, we only need to impose the sign restriction on one of the β_l .

Given the treatment of the default state, the migrations from this state in the composite log-likelihood are taken into account using the estimated constant transition probabilities to states k=1, k=2 and k=3. Furthermore, the composite log-likelihood at lag 2 and its derivative depend on an integral which cannot be computed analytically. This integral is the expected value of

$$\sum_{j=1}^{K} \left[\Phi\left(\frac{c_{k+1} - \delta_j - \beta_j \rho f}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}} \right) - \Phi\left(\frac{c_k - \delta_j - \beta_j \rho f}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}} \right) \right] \times \left[\Phi\left(\frac{c_{j+1} - \delta_l - \beta_l f}{\sigma_l} \right) - \Phi\left(\frac{c_j - \delta_l - \beta_l f}{\sigma_l} \right) \right],$$

where the source of the randomness is f, which follows a standard normal distribution.

Therefore, it can be approximated by

$$\frac{1}{S} \sum_{s=1}^{S} \sum_{j=1}^{K} \left[\left[\Phi \left(\frac{c_{k+1} - \delta_j - \beta_j \rho f_s}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}} \right) - \Phi \left(\frac{c_k - \delta_j - \beta_j \rho f_m}{\sqrt{\sigma_j^2 + \beta_j^2 (1 - \rho^2)}} \right) \right] \times \left[\Phi \left(\frac{c_{j+1} - \delta_l - \beta_l f_s}{\sigma_l} \right) - \Phi \left(\frac{c_j - \delta_l - \beta_l f_s}{\sigma_l} \right) \right],$$

where f_s is simulated S=1,000 times from a normal distribution using $f_s=\rho f_{s-1}+\sqrt{1-\rho^2}\eta_s$, $f_0\sim N(0,1)$ and $\eta_s\sim N(0,1)$, to incorporate the correlation among the factors.

D.2. Results for Design 1

Table 7: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 1 when $\rho=0$

	Mea	n Absolute	e Bias	Standard Errors			
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240	
$c_3 = 1.5$	0.10	0.07	0.05	0.11	0.08	0.06	
$c_4 = 3.0$	0.26	0.17	0.13	0.26	0.19	0.14	
$c_5 = 4.5$	0.45	0.31	0.22	0.43	0.32	0.23	
$c_6 = 6.0$	0.69	0.49	0.35	0.64	0.47	0.34	
$c_7 = 7.5$	0.98	0.69	0.49	0.89	0.64	0.46	
$c_8 = 9.0$	1.30	0.94	0.66	1.20	0.84	0.60	

Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\delta_1 = -0.5$	0.09	0.06	0.05	0.11	0.08	0.06
$\delta_2 = 1.0$	0.11	0.07	0.06	0.13	0.10	0.07
$\delta_3 = 2.5$	0.22	0.15	0.11	0.26	0.19	0.14
$\delta_4 = 4.0$	0.4	0.27	0.20	0.42	0.31	0.23
$\delta_5 = 5.5$	0.63	0.44	0.31	0.63	0.46	0.34
$\delta_6 = 7.0$	0.89	0.64	0.45	0.88	0.64	0.46
$\delta_7 = 8.5$	1.2	0.87	0.62	1.20	0.83	0.59

Unconditional Volatities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\gamma_1 = 1.1$	0.07	0.05	0.03	0.08	0.05	0.04
$\gamma_2 = 1.1$	0.11	0.08	0.05	0.10	0.07	0.05
$\gamma_3 = 1.2$	0.15	0.11	0.08	0.13	0.09	0.07
$\gamma_4 = 1.2$	0.20	0.15	0.10	0.17	0.12	0.08
$\gamma_5 = 1.3$	0.25	0.19	0.13	0.21	0.15	0.10
$\gamma_6 = 1.3$	0.31	0.23	0.16	0.31	0.21	0.15

Table 8: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 1 when $\rho=0.4$

	Mea	n Absolute	e Bias	St	Standard Errors			
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$c_3 = 1.5$	0.11	0.07	0.05	0.13	0.09	0.06		
$c_4 = 3.0$	0.25	0.16	0.12	0.30	0.20	0.15		
$c_5 = 4.5$	0.44	0.28	0.22	0.51	0.34	0.25		
$c_6 = 6.0$	0.67	0.44	0.33	0.77	0.51	0.37		
$c_7 = 7.5$	0.95	0.64	0.47	1.10	0.72	0.51		
$c_8 = 9.0$	1.30	0.87	0.64	1.40	0.97	0.68		
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	$T = 240\ell$		
$\delta_1 = -0.5$	0.16	0.15	0.15	0.14	0.10	0.08		
$\delta_2 = 1.0$	0.13	0.09	0.07	0.17	0.12	0.09		
$\delta_3 = 2.5$	0.23	0.15	0.12	0.30	0.22	0.16		
$\delta_4 = 4.0$	0.39	0.26	0.20	0.50	0.35	0.25		
$\delta_5 = 5.5$	0.61	0.41	0.32	0.75	0.52	0.37		
$\delta_6 = 7.0$	0.89	0.61	0.47	1.10	0.73	0.52		
$\delta_7 = 8.5$	1.20	0.85	0.66	1.40	0.96	0.68		
Unconditional Volatities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\gamma_1 = 1.1$	0.06	0.04	0.03	0.08	0.06	0.04		
$\gamma_2 = 1.1$	0.09	0.07	0.05	0.120	0.08	0.06		
$\gamma_3 = 1.2$	0.14	0.10	0.07	0.16	0.11	0.08		
$\gamma_4 = 1.2$	0.19	0.14	0.10	0.22	0.14	0.10		
$\gamma_5 = 1.3$	0.24	0.17	0.12	0.30	0.18	0.13		
$\gamma_6 = 1.3$	0.30	0.21	0.15	0.50	0.26	0.18		

Table 9: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 1 when $\rho=0.7$

	Mea	n Absolute	e Bias	St	Standard Errors			
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T=240		
$c_3 = 1.5$	0.13	0.08	0.07	0.21	0.13	0.07		
$c_4 = 3.0$	0.29	0.19	0.15	0.52	0.32	0.16		
$c_5 = 4.5$	0.50	0.33	0.25	0.93	0.58	0.28		
$c_6 = 6.0$	0.76	0.50	0.38	1.40	0.92	0.43		
$c_7 = 7.5$	1.10	0.72	0.53	2.00	1.30	0.61		
$c_8 = 9.0$	1.50	0.98	0.72	2.40	1.70	0.79		
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\delta_1 = -0.5$	0.29	0.30	0.31	0.16	0.13	0.096		
$\delta_2 = 1.0$	0.16	0.11	0.08	0.22	0.16	0.10		
$\delta_3 = 2.5$	0.27	0.18	0.14	0.52	0.32	0.18		
$\delta_4 = 4.0$	0.46	0.31	0.24	0.87	0.58	0.29		
$\delta_5 = 5.5$	0.72	0.49	0.39	1.40	0.93	0.44		
$\delta_6 = 7.0$	1.10	0.72	0.57	2.10	1.40	0.63		
$\delta_7 = 8.5$	1.50	1.00	0.82	2.30	1.60	0.78		
Unconditional Volatities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\gamma_1 = 1.1$	0.05	0.04	0.03	0.17	0.09	0.04		
$\gamma_2 = 1.1$	0.10	0.07	0.05	0.24	0.15	0.06		
$\gamma_3 = 1.2$	0.15	0.10	0.08	0.34	0.22	0.09		
$\gamma_4 = 1.2$	0.20	0.14	0.10	0.43	0.29	0.12		
$\gamma_5 = 1.3$	0.26	0.18	0.13	0.72	0.46	0.16		
$\gamma_6 = 1.3$	0.31	0.22	0.15	0.49	0.35	0.20		

Figure 3: Empirical PDF of t-Statistic for c_{k+1} when $\rho = 0$

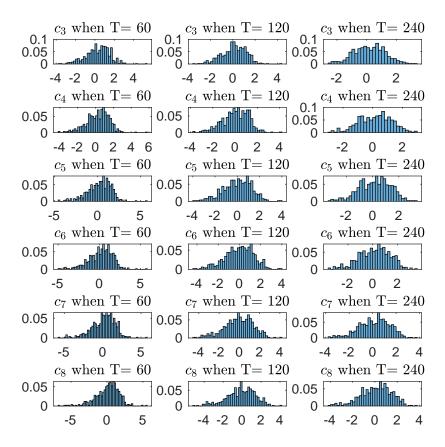


Figure 4: Empirical PDF of t-Statistic for c_{k+1} when $\rho = 0.4$

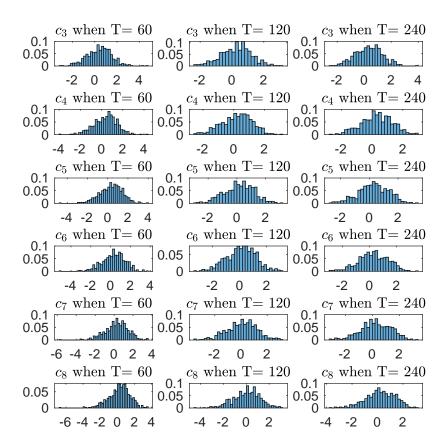


Figure 5: Empirical PDF of t-Statistic for c_{k+1} when $\rho = 0.7$

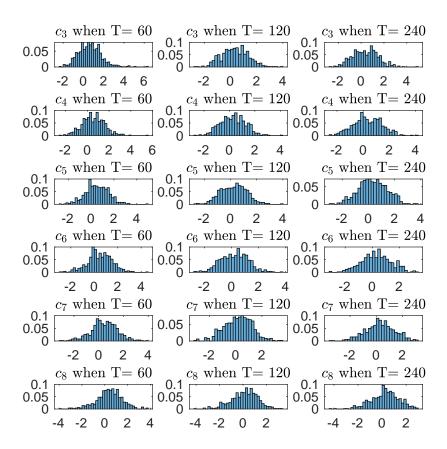


Figure 6: Empirical PDF of t Statistic for δ_l when $\rho = 0$

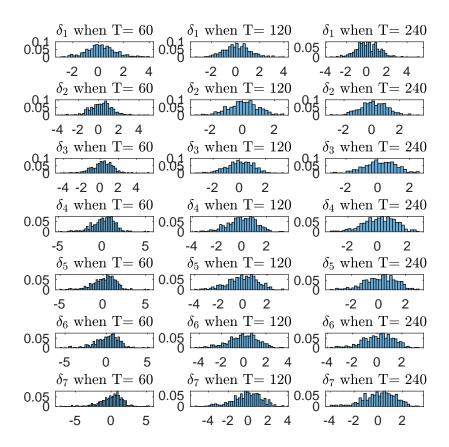


Figure 7: Empirical PDF of t-Statistic for δ_l when $\rho = 0.4$

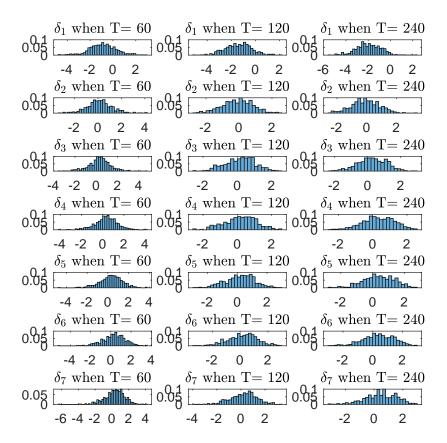


Figure 8: Empirical PDF of t-Statistic for δ_l when $\rho=0.7$

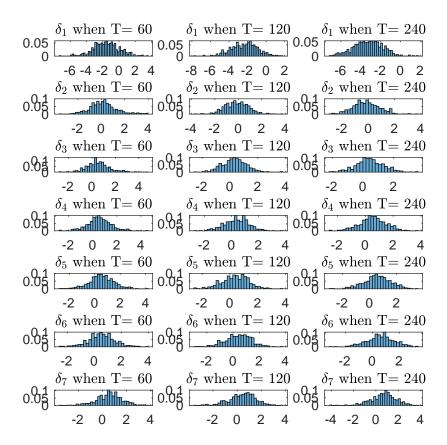


Figure 9: Empirical PDF of t-Statistic for γ_l when $\rho = 0$

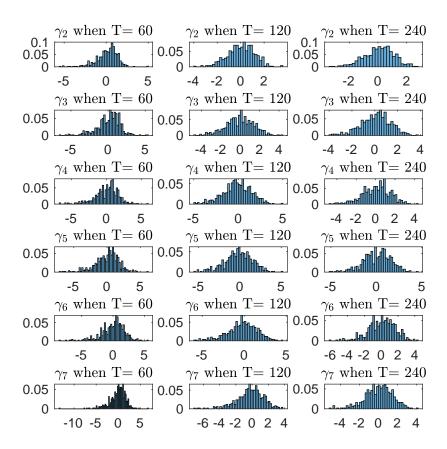


Figure 10: Empirical PDF of t-Statistic for γ_l when $\rho = 0.4$

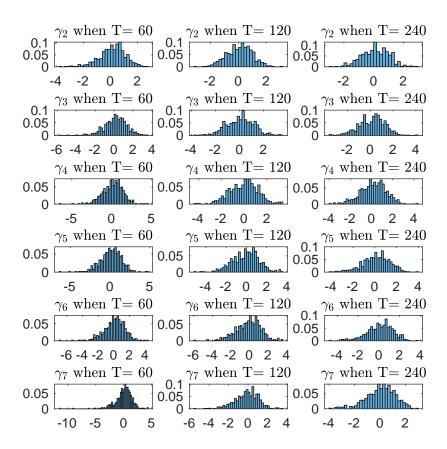
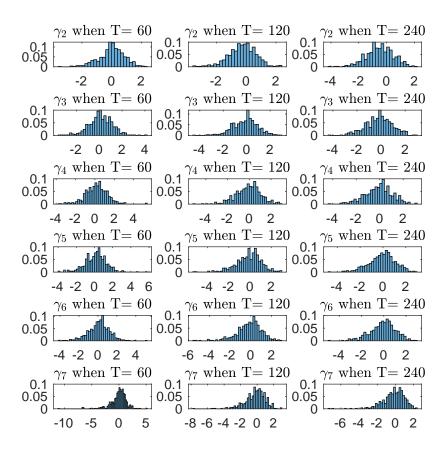


Figure 11: Empirical PDF of t-Statistic for γ_l when $\rho = 0.7$



D.3. Results for Design 2

Table 10: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 2 when $\rho=0$

	Mea	n Absolute	Bias	St	andard Eri	rors
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T=240
$c_3 = 1.5$	0.09	0.07	0.05	0.18	0.11	0.08
$c_4 = 3.0$	0.15	0.11	0.08	0.33	0.21	0.14
$c_5 = 4.5$	0.16	0.11	0.08	0.50	0.32	0.23
$c_6 = 6.0$	0.16	0.12	0.091	0.72	0.50	0.36
$c_7 = 7.5$	0.25	0.20	0.16	1.00	0.74	0.55
$c_8 = 9.0$	0.45	0.35	0.30	1.40	1.10	0.79
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\delta_1 = -0.5$	0.09	0.07	0.05	0.13	0.09	0.07
$\delta_2 = 1.0$	0.10	0.08	0.06	0.17	0.12	0.08
$\delta_3 = 2.5$	0.14	0.11	0.08	0.31	0.19	0.14
$\delta_4 = 4.0$	0.16	0.12	0.09	0.47	0.31	0.22
$\delta_5 = 5.5$	0.17	0.12	0.09	0.68	0.47	0.34
$\delta_6 = 7.0$	0.24	0.19	0.16	0.96	0.69	0.51
$\delta_7 = 8.5$	0.41	0.32	0.27	1.30	0.99	0.74
Factor Sensitivities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\beta_2 = 0.74$	0.21	0.19	0.19	1.70	1.30	0.86
$\beta_3 = 0.78$	0.15	0.14	0.12	1.40	1.10	0.77
$\beta_4 = 0.82$	0.15	0.14	0.12	1.60	1.20	0.98
$\beta_5 = 0.86$	0.17	0.15	0.13	1.60	1.10	0.87
$\beta_6 = 0.9$	0.21	0.18	0.16	1.50	1.00	0.65
$\beta_7 = 0.95$	0.32	0.34	0.40	1.30	0.85	0.60
Volatilities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\sigma_1 = 0.71$	0.20	0.20	0.18	1.50	0.92	0.68
$\sigma_2 = 0.74$	0.13	0.12	0.13	1.30	0.93	0.62
$\sigma_3 = 0.78$	0.14	0.12	0.11	1.20	0.87	0.64
$\sigma_4 = 0.82$	0.16	0.15	0.12	1.30	1.00	0.82
$\sigma_5 = 0.86$	0.17	0.16	0.13	1.30	0.98	0.74
$\sigma_6 = 0.9$	0.24	0.20	0.20	1.30	0.84	0.58
$\sigma_7 = 0.95$	0.27	0.26	0.28	0.95	0.58	0.34
Autocorrelation	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\rho = 0$	1.9e - 3	1.9e - 3	1.7e - 3	0.19	0.14	0.11

Table 11: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 2 when $\rho=0.4$

	Mea	n Absolute	e Bias	Standard Errors					
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240			
$c_3 = 1.5$	0.15	0.12	0.11	0.34	0.24	0.17			
$c_4 = 3.0$	0.23	0.18	0.18	0.63	0.45	0.32			
$c_5 = 4.5$	0.24	0.18	0.17	0.92	0.66	0.48			
$c_6 = 6.0$	0.22	0.17	0.13	1.20	0.90	0.67			
$c_7 = 7.5$	0.27	0.22	0.15	1.60	1.20	0.95			
$c_8 = 9.0$	0.49	0.40	0.34	2.10	1.60	1.30			
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240			
$\delta_1 = -0.5$	0.13	0.11	0.09	0.17	0.12	0.09			
$\delta_2 = 1.0$	0.14	0.11	0.09	0.30	0.22	0.16			
$\delta_3 = 2.5$	0.21	0.17	0.15	0.59	0.44	0.31			
$\delta_4 = 4.0$	0.23	0.18	0.16	0.89	0.65	0.46			
$\delta_5 = 5.5$	0.23	0.17	0.14	1.20	0.89	0.66			
$\delta_6 = 7.0$	0.28	0.23	0.17	1.60	1.20	0.93			
$\delta_7 = 8.5$	0.52	0.43	0.38	2.10	1.60	1.30			
Factor Sensitivities			T = 240	T = 60	T = 120	T = 240			
$\beta_2 = 0.77$	0.18	0.16	0.15	0.96	0.67	0.45			
$\beta_3 = 0.81$	0.15	0.12	0.10	0.87	0.64	0.42			
$\beta_4 = 0.85$	0.16	0.13	0.10	0.93	0.62	0.45			
$\beta_5 = 0.9$	0.19	0.15	0.12	0.99	0.67	0.42			
$\beta_6 = 0.94$	0.23	0.20	0.16	1.10	0.82	0.44			
$\beta_7 = 0.99$	0.33	0.29	0.27	1.10	0.82	0.50			
Volatilities	I .		T = 240	T = 60	T = 120	T = 240			
$\sigma_1 = 0.74$	0.18	0.17	0.16	1.30	0.79	0.56			
$\sigma_2 = 0.77$	0.15	0.15	0.12	0.91	0.61	0.41			
$\sigma_3 = 0.81$	0.17	0.15	0.11	0.82	0.63	0.43			
$\sigma_4 = 0.85$	0.2	0.17	0.11	0.91	0.63	0.46			
$\sigma_5 = 0.9$	0.25	0.23	0.13	0.88	0.70	0.46			
$\sigma_6 = 0.94$	0.28	0.27	0.14	1.10	0.85	0.52			
$\sigma_7 = 0.99$	0.45	0.46	0.23	1.10	0.73	0.55			
Autocorrelation	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240			
$\rho = 0.4$	0.16	0.14	0.11	0.60	0.44	0.26			

Table 12: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 2 when $\rho=0.7$

	Mea	n Absolute	e Bias	Standard Errors		
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$c_3 = 1.5$	0.19	0.15	0.12	0.61	0.34	0.21
$c_4 = 3.0$	0.31	0.23	0.19	1.10	0.63	0.39
$c_5 = 4.5$	0.33	0.25	0.20	1.50	0.90	0.55
$c_6 = 6.0$	0.33	0.25	0.19	2.00	1.20	0.74
$c_7 = 7.5$	0.40	0.31	0.22	2.40	1.50	0.97
$c_8 = 9.0$	0.62	0.48	0.34	3.00	2.00	1.30
	I			I		
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\delta_1 = -0.5$	0.25	0.21	0.20	0.23	0.17	0.13
$\delta_2 = 1.0$	0.19	0.15	0.11	0.39	0.27	0.18
$\delta_3 = 2.5$	0.28	0.24	0.20	0.94	0.60	0.38
$\delta_4 = 4.0$	0.31	0.27	0.23	1.40	0.90	0.57
$\delta_5 = 5.5$	0.34	0.29	0.27	1.90	1.20	0.75
$\delta_6 = 7.0$	0.47	0.41	0.37	2.40	1.50	0.98
$\delta_7 = 8.5$	0.81	0.71	0.63	3.00	2.00	1.30
Factor Sensitivities			T = 240	T = 60	T = 120	T = 240
$\beta_2 = 0.85$	0.23	0.19	0.18	0.86	0.60	0.46
$\beta_3 = 0.9$	0.19	0.15	0.14	0.88	0.62	0.46
$\beta_4 = 0.94$	0.18	0.16	0.13	0.95	0.65	0.49
$\beta_5 = 0.99$	0.23	0.20	0.18	0.99	0.67	0.49
$\beta_6 = 1.0$	0.29	0.25	0.20	1.20	0.86	0.52
$\beta_7 = 1.1$	0.37	0.32	0.26	1.20	0.87	0.49
Volatilities	I .		T = 240	T = 60	T = 120	T = 240
$\sigma_1 = 0.81$	0.21	0.20	0.23	1.40	1.10	1.00
$\sigma_2 = 0.85$	0.20	0.17	0.15	0.82	0.66	0.48
$\sigma_3 = 0.9$	0.19	0.17	0.15	0.86	0.69	0.51
$\sigma_4 = 0.94$	0.19	0.17	0.14	1.30	0.67	0.53
$\sigma_5 = 0.99$	0.21	0.19	0.17	1.00	0.82	0.56
$\sigma_6 = 1.0$	0.28	0.25	0.19	1.20	1.10	0.73
$\sigma_7 = 1.1$	0.40	0.33	0.29	1.30	1.10	0.75
Autocorrelation	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
$\rho = 0.7$	0.31	0.29	0.31	0.62	0.52	0.35

Table 13: Average of Estimated Downgrade Probabilities and Probabilities of Default (in %) for Design 2

$\rho = 0$	True Value	T = 60	T = 120	T = 240
$\overline{DP(1 A)}$	32.52	33.29	33.48	33.55
$DP\left(2 A\right)$	43.35	44.67	44.85	44.87
PD(1 A)	1.4e - 7	1.4e - 5	3.7e - 6	1.5e - 6
PD(12 A)	3.28	3.57	3.61	3.61
PD(24 A)	3.19	3.16	3.19	3.19
PD(36 A)	2.91	3.14	3.17	3.17
$\rho = 0.4$	True Value	T = 60	T = 120	T = 240
$\overline{DP(1 A)}$	32.44	34.8	35.1	34.33
DP(2 A)	44.05	45.89	46.37	46.33
PD(1 A)	8.2e - 8	2.2e - 4	6.3e - 5	1.4e - 5
PD(12 A)	3.37	3.81	3.88	3.93
PD(24 A)	3.68	3.46	3.62	3.67
PD(36 A)	3.16	3.44	3.61	3.66
$\rho = 0.7$	True Value	T = 60	T = 120	T = 240
$\overline{DP(1 A)}$	34.7	37.69	38.21	38.48
$DP\left(2 A\right)$	44.93	48.23	48.95	49.3
PD(1 A)	1.7e - 5	2.0e - 3	8.6e - 4	3.4e - 4
PD(12 A)	3.77	4.26	4.41	4.55
PD(24 A)	4.53	4.03	4.18	4.33
PD(36 A)	3.74	4.01	4.16	4.31

D.4. Results for Design 3

Table 14: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 3 when $\rho=0$

-	Mea	n Absolute	Bias	St	Standard Errors			
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T=240		
$c_3 = 1.5$	0.08	0.06	0.05	0.14	0.10	0.07		
$c_4 = 3.0$	0.12	0.09	0.07	0.27	0.19	0.14		
$c_5 = 4.5$	0.13	0.10	0.07	0.43	0.30	0.23		
$c_6 = 6.0$	0.13	0.09	0.07	0.64	0.47	0.36		
$c_7 = 7.5$	0.20	0.15	0.13	0.92	0.69	0.53		
$c_8 = 9.0$	0.34	0.27	0.23	1.20	0.95	0.73		
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\delta_1 = -0.5$	0.09	0.07	0.05	0.12	0.09	0.06		
$\delta_2 = 1.0$	0.10	0.07	0.06	0.15	0.10	0.07		
$\delta_3 = 2.5$	0.13	0.09	0.07	0.25	0.18	0.13		
$\delta_4 = 4.0$	0.14	0.10	0.07	0.4	0.28	0.21		
$\delta_5 = 5.5$	0.14	0.10	0.08	0.60	0.43	0.33		
$\delta_6 = 7.0$	0.19	0.15	0.13	0.85	0.63	0.48		
$\delta_7 = 8.5$	0.31	0.24	0.21	1.20	0.88	0.68		
Factor Sensitivities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\beta_2 = 0.71$	0.21	0.18	0.19	1.60	1.10	0.80		
$\beta_3 = 071$	0.13	0.11	0.11	1.20	0.85	0.65		
$\beta_4 = 0.71$	0.14	0.12	0.11	1.30	0.96	0.73		
$\beta_5 = 0.71$	0.14	0.12	0.12	1.10	0.87	0.62		
$\beta_6 = 0.71$	0.16	0.14	0.13	0.90	0.68	0.45		
$\beta_7 = 0.71$	0.24	0.26	0.30	0.87	0.56	0.38		
Volatilities	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\sigma_1 = 0.71$	0.19	0.19	0.19	1.20	0.85	0.61		
$\sigma_2 = 0.74$	0.13	0.12	0.13	1.10	0.76	0.54		
$\sigma_3 = 0.78$	0.12	0.10	0.09	0.91	0.67	0.50		
$\sigma_4 = 0.82$	0.12	0.11	0.10	0.93	0.72	0.55		
$\sigma_5 = 0.86$	0.14	0.11	0.10	0.82	0.63	0.46		
$\sigma_6 = 0.9$	0.17	0.13	0.13	1.10	0.48	0.33		
$\sigma_7 = 0.95$	0.19	0.18	0.19	0.52	0.31	0.21		
Autocorrelation	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\rho = 0.0$	1.7e - 3	1.5e - 3	1.5e - 3	0.17	0.13	0.09		

Table 15: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 3 when $\rho=0.4$

	Mea	n Absolute	e Bias	St	Standard Errors		
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T=240	
$c_3 = 1.5$	0.14	0.11	0.09	0.32	0.21	0.15	
$c_4 = 3.0$	0.2	0.15	0.13	0.59	0.40	0.27	
$c_5 = 4.5$	0.20	0.15	0.12	0.85	0.59	0.40	
$c_6 = 6.0$	0.18	0.13	0.09	1.20	0.82	0.56	
$c_7 = 7.5$	0.22	0.18	0.13	1.50	1.10	0.77	
$c_8 = 9.0$	0.40	0.33	0.26	2.00	1.40	1.00	
	ı						
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240	
$\delta_1 = -0.5$	0.13	0.10	0.09	0.16	0.12	0.09	
$\delta_2 = 1.0$	0.14	0.11	0.08	0.28	0.20	0.14	
$\delta_3 = 2.5$	0.19	0.15	0.12	0.55	0.39	0.26	
$\delta_4 = 4.0$	0.20	0.15	0.12	0.83	0.58	0.40	
$\delta_5 = 5.5$	0.19	0.14	0.10	1.10	0.80	0.55	
$\delta_6 = 7.0$	0.23	0.18	0.14	1.50	1.10	0.74	
$\delta_7 = 8.5$	0.40	0.32	0.27	1.90	1.40	0.99	
Factor Sensitivities			T = 240	T = 60	T = 120		
$\beta_2 = 0.74$	0.16	0.14	0.12	0.77	0.61	0.45	
$\beta_3 = 074$	0.13	0.10	0.09	0.67	0.59	0.43	
$\beta_4 = 0.74$	0.14	0.11	0.09	0.67	0.58	0.46	
$\beta_5 = 0.74$	0.15	0.12	0.10	0.65	0.54	0.41	
$\beta_6 = 0.74$	0.17	0.14	0.11	0.69	0.56	0.41	
$\beta_7 = 0.74$	0.20	0.19	0.20	0.73	0.48	0.33	
Volatilities	l		T = 240	T = 60	T = 120	T = 240	
$\sigma_1 = 0.74$	0.19	0.16	0.14	1.00	0.69	0.54	
$\sigma_2 = 0.77$	0.13	0.11	0.09	0.64	0.52	0.40	
$\sigma_3 = 0.81$	0.12	0.10	0.09	0.60	0.51	0.40	
$\sigma_4 = 0.85$	0.13	0.11	0.09	0.59	0.48	0.38	
$\sigma_5 = 0.9$	0.13	0.13	0.10	0.53	0.43	0.33	
$\sigma_6 = 0.94$	0.16	0.15	0.12	0.55	0.45	0.35	
$\sigma_7 = 0.99$	0.23	0.22	0.21	0.55	0.37	0.27	
Autocorrelation	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240	
$\rho = 0.4$	0.14	0.12	0.12	0.57	0.43	0.34	

Table 16: Average of Absolute Bias and Standard Errors for Estimated Parameters for Design 3 when $\rho=0.7$

	Mea	n Absolute	e Bias	St	Standard Errors			
Thresholds	T = 60	T = 120	T = 240	T = 60	T = 120	T=240		
$c_3 = 1.5$	0.17	0.13	0.11	0.48	0.28	0.19		
$c_4 = 3.0$	0.26	0.21	0.17	0.89	0.54	0.36		
$c_5 = 4.5$	0.28	0.22	0.17	1.30o	0.78	0.52		
$c_6 = 6.0$	0.28	0.22	0.17	1.60	1.10	0.71		
$c_7 = 7.5$	0.36	0.27	0.20	2.10	1.40	0.95		
$c_8 = 9.0$	0.57	0.44	0.33	2.60	1.80	1.20		
	ı							
Intercepts	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\delta_1 = -0.5$	0.24	0.21	0.20	0.22	0.18	0.13		
$\delta_2 = 1.0$	0.19	0.14	0.10	0.36	0.25	0.17		
$\delta_3 = 2.5$	0.25	0.21	0.17	0.8	0.53	0.36		
$\delta_4 = 4.0$	0.28	0.23	0.20	1.20	0.79	0.54		
$\delta_5 = 5.5$	0.3	0.25	0.22	1.60	1.10	0.71		
$\delta_6 = 7.0$	0.42	0.34	0.31	2.00	1.40	0.93		
$\delta_7 = 8.5$	0.68	0.56	0.50	2.60	1.80	1.20		
Factor Sensitivities			T = 240	T = 60	T = 120	T = 240		
$\beta_2 = 0.81$	0.22	0.18	0.16	0.86	0.62	0.56		
$\beta_3 = 0.81$	0.17	0.14	0.12	0.86	0.68	0.59		
$\beta_4 = 0.81$	0.18	0.14	0.13	0.89	0.65	0.54		
$\beta_5 = 0.81$	0.23	0.18	0.17	0.95	0.76	0.54		
$\beta_6 = 0.81$	0.23	0.19	0.18	0.98	0.80	0.53		
$\beta_7 = 0.81$	0.28	0.22	0.18	0.78	0.60	0.38		
Volatilities	l		T = 240	T = 60	T = 120	T = 240		
$\sigma_1 = 0.81$	0.21	0.21	0.25	1.70	1.10	1.20		
$\sigma_2 = 0.84$	0.18	0.14	0.14	0.81	0.61	0.56		
$\sigma_3 = 0.90$	0.18	0.15	0.14	1.10	0.66	0.55		
$\sigma_4 = 0.94$	0.18	0.14	0.13	0.75	0.60	0.49		
$\sigma_5 = 0.99$	0.21	0.19	0.16	0.81	0.69	0.51		
$\sigma_6 = 1.0$	0.23	0.21	0.19	0.81	0.78	0.56		
$\sigma_7 = 1.1$	0.33	0.29	0.24	0.88	0.59	0.48		
Autocorrelation	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240		
$\rho = 0.7$	0.32	0.31	0.31	0.61	0.56	0.40		

Table 17: Average of Estimated Downgrade Probabilities and Probabilities of Default (in %) for Design 3

$\rho = 0$	True Value	T = 60	T = 120	T = 240
$\overline{DP(1 A)}$	31.75	32.5	32.72	32.71
$DP\left(2 A\right)$	43.27	44.53	44.74	44.67
PD(1 A)	2.5e - 8	3.4e - 6	1.0e - 6	2.7e - 7
PD(12 A)	3.26	3.57	3.58	3.58
PD(24 A)	3.14	3.14	3.16	3.16
PD(36 A)	2.91	3.12	3.14	3.13
$\rho = 0.4$	True Value	T = 60	T = 120	T=240
$\overline{DP(1 A)}$	32.45	33.94	34.25	34.26
DP(2 A)	43.91	45.67	46.04	46.06
PD(1 A)	1.2e - 7	3.5e - 5	8.0e - 6	2.6e - 6
PD(12 A)	3.34	3.76	3.82	3.83
PD(24 A)	3.58	3.35	3.43	3.48
PD(36 A)	3.11	3.34	3.42	3.47
$\rho = 0.7$	True Value	T = 60	T = 120	T = 240
DP(1 A)	34.01	36.85	37.31	37.5
DP(2 A)	44.69	47.73	48.36	48.61
PD(1 A)	4.5e - 6	7.4e - 4	1.5e - 4	9.3e - 5
PD(12 A)	3.60	4.10	4.22	4.31
PD(24 A)	4.337	3.82	3.89	3.90
PD(36 A)	3.55	3.80	3.87	3.88

References

- ALTMAN, E., and A., SAUNDERS (2004). Credit Risk Measurement: Developments Over the Last 20 Years. Journal of Banking & Finance, 21, 1721–1742.
- AZIZPOUR, S., GIESECKE, K., and G., SCHWENKLER (2018). Exploring the Sources of Default Clustering, Journal of Financial Economics, 129, 154–183.
- Basel Committee on Banking Supervision (2004). International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Technical report, Bank for International Settlements.
- Basel Committee on Banking Supervision (2009). Guidelines for Computing Capital for Incremental Risk in the Trading Book, Technical report, Bank for International Settlements.
- BERNDT, A., DOUGLAS, R., DUFFIE, D., and M., FERGUSSON (2018). Corporate Credit Risk Premia, Review of Finance, 419–454.
- Bonhomme, S., Jochmans, K., and J., Robin (2017). Nonparametric Estimation of Non-Exchangeable Latent-Variable Models, Journal of Econometrics, 201, 237–248.
- Cox, D., and N., Reid (2004). A Note on Pseudolikelihood Constructed from Marginal Densities. Biometrika, 91, 729-737.
- CREAL, D., KOOPMAN, S., and A., LUCAS (2012). Generalized Autoregressive Score Models with Applications, Journal of Applied Econometrics, 28, 777–795.
- CREAL, D., SCHWAAB, B., KOOPMAN, S., and A., Lucas (2014). Observation-Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk, Review of Economics and Statistics, 96, 898–915.
- CSORGO, M. (1967). On the Strong Law of Large Numbers and the Central Limit Theorem for Martingales. Transactions of the American Mathematical Society, 259–275.
- Duffie, D., Eckner, A., Horel, G., and L., Saita (2009). Frailty Correlated Default, The Journal of Finance, 64,2089–2123.
- FARMER, L. (2021). The Discretization Filter: A Simple Way to Estimate Nonlinear State Space Models, Quantitative Economics, 12, 41-76.
- FENG, D., GOURIÉROUX, C., and J., JASIAK (2008). The Ordered Qualitative Model for Credit Rating Transitions, Journal of Empirical Finance, 15, 111–130.
- Fuh, C. (2006). Efficient Likelihood Estimation in State Space Models, Annals of Statistics, 34, 2026-2068.

- GAGLIARDINI, P., and C., GOURIÉROUX (2005). Migration Correlation: Definition and Efficient Estimation, Journal of Banking and Finance, 29, 865–894.
- GAGLIARDINI, P., and C., GOURIÉROUX (2014). Efficiency in Large Dynamic Panel Models with Common Factors, Econometric Theory, 30, 961–1020.
- Gagliardini, P., and C., Gouriéroux (2015). Granularity Theory with Applications to Finance and Insurance, Cambridge University Press, 186 pages.
- Gouriéroux, C., and A., Monfort (2018). Composite Indirect Inference with Application to Corporate Risks, Econometrics and Statistics, 7, 30–45.
- Gouriéroux, C., Monfort, A., and V., Polimenis (2006). Affine Models for Credit Risk Analysis, Journal of Financial Econometrics, 4, 494-530.
- Gouriéroux, C., Monfort, A., Mouabbi, S., and J.P., Renne (2021). *Disastrous Defaults*, Review of Finance, forthcoming.
- Hajek, J., and A., Renyi (1955). Generalization of an Inequality of Kolmogorov, Acta Math Acad. Sci. Hungar., 6, 281–283.
- HARKO, T., LOBO, F., and M., MAK(2016). Exact Analytical Solutions of the Susceptible-Infected-Recovered (SIR) Epidemic Model and of the SIR Model with Equal Death and Birth Rates, Applied Mathematics and Computation, 236, 184–194.
- Hubert, P.(1967). The Behavior of Maximum Likelihood Estimation under Nonstandard Conditions, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Le Cam, L., and Neyman, J, eds, University of California Press, 221–233.
- Hull, J. (2012). Risk Management and Financial Institutions, John Wiley & Sons, 733 pages.
- JENNRICH, R. I.(1969). Asymptotic Properties of Non-Linear Least Squares Estimators, Ann. Math. Statist., 40, 633–643.
- Koopman, S., Lucas, A., and A., Monteiro (2008). The Multi-State Latent Factor Intensity Model for Credit Rating Transitions, Journal of Econometrics, 142, 399–424.
- LINDSAY, B. G.(1988). Composite Likelihood Methods, Contemporary Mathematics, 80, 221–239.
- NEWEY, W., and K., West (1994). Automatic Lag Selection in Covariance Matrix Estimation, The Review of Economic Studies, 61(4), 631–653.
- NICKELL, P., PERRAUDIN, W, and S., VAROTTO (2001). Stability of Rating Transitions, Journal of Banking and Finance, 24, 203–227.

- Tuzcuoglu, K. (2019). Composite Likelihood Estimation of an Autoregressive Panel Probit Model with Random Effects. Bank of Canada, DP 2019–06.
- VARIAN, C.(2008). On Composite Marginal Likelihoods, AStA Advances in Statistical Analysis, Springer; German Statistical Society, 92(1), 1–28.
- Varian, C., Reid, N., and D., Firth (2011). An Overview of Composite Likelihood Methods. Statistica Sinica, 5–42.
- Vasicek, O. (1991). Limiting Loan Loss Probability Distribution. DP KMV Corporation.
- VASICEK, O. (2015). Probability Loss and Loan Portfolio. In: Vasicek, O.A. (ed.) Finance, Economics and Mathematics. Hoboken, NJ: John Wiley & Sons, Inc., Chapter 17.
- WHITE, H. (1982). Maximum Likelihood Estimation of Misspecified Models. Econometrica, 50, 1–25.