Identification of hidden Markov chains governing dependent credit-rating migrations*

D.V. Boreiko[1], S.Y. Kaniovski[2], Y.M. Kaniovski[3] and G.Ch. Pflug.[4]

Abstract

Three models of dependent credit-rating migrations are considered. Each of them entails a coupling scheme and a discrete-time Markovian macroeconomic dynamics. Every credit-rating migration is modeled as a mixture of an idiosyncratic and a common component. The larger is the pool of debtors affected by the same common component, the stronger is the dependence among migrations. The distribution of the common component depends on macroeconomic conditions. At every time instant, the resulting allocation of debtors to credit classes and industries follows a mixture of multinomial distributions.

Dealing with M non-default credit-classes, there are 2^M theoretically possible macroeconomic outcomes. Only few of them occur with a positive probability. Restricting the macroeconomic dynamics to such outcomes simplifies estimation. A heuristics for identifying them is suggested. Using the maximum likelihood method, it was tested on a Standard and Poor's (S&P's) dataset.

Keywords: Markov chain, mixture, maximum likelihood, multinomial distribution, heuristics.

Math. Subj. Class. (2000): 90C30, 90C90

*The authors acknowledge financial support from the Free University of Bozen-Bolzano (WW 2086).

[1]Faculty of Economics and Management, Free University of Bozen-Bolzano, piazza Università 1, I-39100, Bolzano, Italy.
E-mail: DBoreiko@unibz.it. Corresponding author. Tel. +39 0471 013 277.

[2]Macroeconomics and European Economic Policy Research Group, Austrian Institute for Economic Research (WIFO), Arsenal 20, A-1030 Vienna, Austria.
E-mail: Serguei.Kaniovski@wifo.ac.at.

[3]Faculty of Economics and Management, Free University of Bozen-Bolzano, piazza Università 1, I-39100 Bolzano, Italy.
E-mail: YKaniovskyi@unibz.it.

[4]Department of Statistics and Decision Support Systems, University of Vienna, Universitätstraße 5, A-1090 Vienna, Austria.
E-mail: georg.pflug@univie.ac.at.

1 Introduction

CreditMetrics, the first available portfolio model for evaluating credit risk, is based on credit migration analysis. It involves estimating the probability of moving from one credit quality to another within a given horizon. The model was created by the JP Morgan bank in 1997. See Gupton et al. (1997). Since 1999, it became part of the credit risk management for almost all major banks.

Within this model, migrations of a debtor are treated as a trajectory of a time-homogeneous Markov chain. Its transition matrix P is estimated as the time average of migration counts reported by a rating agency. The corresponding transition probabilities are referred two as historical or unconditional. The adjective "historical" refers to the fact that P is based on a long period of observation. (Under the assumption of stationarity, a longer period of observation implies a more precise estimate of P.) The adjective "unconditional" means that macroeconomic factors, like recession or expansion, are not explicitly taken into consideration. Integrating the state of the economy into modeling, Bangia et al. (2002), McNeil and Wendin (2007), Korolkiewicz and Elliott (2008), Frydman and Schuermann (2008), Fei et al. (2012), Xing et al. (2012) extended this unconditional view to a conditional perspective. Some of the models, entail a regime-switching matrix as a means for modeling of macroeconomic dynamics. For example, analyzing data of the National Bureau of Economic Research (NBER), Bangia et al. observe that that: "While the 1981–1998 regime-switching matrix implies that on average 17.8% of all quarters are contraction quarters, the 1959–1998 regime-switching matrix predicts 20.9% of all quarters to be contractions, indicating that the economic development over the last 20 years has been relatively favorable. Moreover, recessions seem to be getting shorter as evidenced by the maintain probability declining from 42.4% (1959–1998) to 30.8% (1981–1998)". (See p. 467.) Their quarterly 2×2 regime-switching matrix contains probabilities of the events that a recession (expansion) now will be followed by a recession (expansion) in the next quarter of a year. This is a probabilistic description of business cycles in the US economy.

Kaniovski and Pflug (2007) suggested a coupling scheme for rendering individual credit-rating migrations dependent. The term coupling scheme means that every migration is a weighted sum of two components, each governed by the same historical matrix P: an idiosyncratic and a common one. The weights are realizations of Bernoulli random variables. Their probabilities of success determine the frequency of idiosyncratic migrations among the corresponding debtors in the pool. In the simplest case, the probabilities are specific for each credit class. In a more realistic setting, the probabilities depend on credit class and industry. The conditional distribution of a common component depends on a binary tendency variable. Such a variable is assigned to each non-default credit class. The value 0 (1) implies an increased (smaller) probability of worsening of the credit quality. Probabilities of migration towards more secure credit ratings are adjusted in the opposite direction: they are higher (lower) if the value 1 (0) is taken. Considering M non-default credit classes, a binary M-vector $\vec{\chi}$ is necessary to define the conditional migration probabilities of every debtor. Its i -th coordinate χ_i is the binary variable assigned to the credit class i. All other things equal, dependence among migrations is stronger, if the pool of debtors affected by the same common component is bigger. In this sense, the model of

Kaniovski and Pflug (2007), where migrations of all debtors belonging to a credit class entail the same common component, is characterized by the strongest dependence among migrations. (Note that a credit class is the next largest, after the whole portfolio, classification group.) A weaker dependence pattern exhibits the coupling scheme suggested in Wozabal and Hochreiter (2012). In this case, common components are conditionally independent across debtors. (Because the conditional migration probabilities of each debtor from a credit class i involve χ_i, migrations of such debtors are stochastically dependent.) The model introduced in Boreiko et. al (2016), assigns a common component to every combination of a non-default credit class and an industry. Since the same common component affects here a smaller (larger) pool of debtors, dependence among migrations is weaker (stronger) than under the assumptions of Kaniovski and Pflug (2007) (Wozabal and Hochreiter (2012)).

Originally, coupling schemes were used for modeling of dependent credit-rating migrations and losses generated by a portfolio. As a result of simulations, Kaniovski and Pflug (2007) report cascades of defaults and a loss distribution with a heavy tail. Estimating parameters of coupling schemes from historical migrations, it became clear that tendency variables can be interpreted as macroeconomic factors. A business cycle entails two phases: a recession and an expansion. A recession is characterized by higher (lower) downgrading (upgrading) probabilities. For an expansion, the probabilities are adjusted in the opposite direction. Considering one non-default credit class and associating the value 1 (0) with an expansion (recession), the impact of a business cycle on credit rating migrations can be encoded by a tendency variable. Introducing a dynamics of tendency variables, renders coupling schemes a means for modeling macroeconomic dynamics.

Modeling migrations of debtors belonging to M non-default credit classes and S industries, there are $M \cdot S + 2^M$ parameters defining each of the three coupling schemes. Out of these values, $M \cdot S$ are the probabilities of success of Bernoulli random variables determining the frequencies of idiosyncratic migrations and 2^M are the probabilities π_i assigned to the binary vectors $\vec{\chi}^{(i)}$. (Since every coordinate takes on two values, there are altogether 2^M binary M-vectors. To refer to them correctly, they have to be numbered.)

In this paper, a generalization involving a Markovian dynamics of tendency variables is considered for each of the known coupling schemes. The corresponding maximum likelihood estimators turn out to be a hard problem: dealing with M non-default credit classes and S industries, there are $M \cdot S + 2^M + 2^{2M}$ parameters. This is a huge number of unknowns given that in many applications $M = 7$. The term 2^{2M} corresponds to the number of entries of a Markovian $2^M \times 2^M$ matrix $\mathcal{P}$ governing the dynamics of tendency variables. Its entry $p_{i,j}$ equals the probability that a binary vector $\vec{\chi}^{(i)}$ is followed by a binary vector $\vec{\chi}^{(j)}$. The majority of the binary vectors are attained with a vanishing probability. Consequently, the corresponding macroeconomic scenarios can be ignored in the risk analysis.* The heuristic suggested here exploits this fact: instead of all 2^M binary vectors, a subset involving only significant, in certain sense, macroeconomic outcomes is

*The banking supervision Accords (recommendations on banking regulations) – Basel I, Basel II and Basel III – issued by the Basel Committee on Banking Supervision (BCBS) set typically 0.1% as the upper limit for the probability of undesirable credit events.

considered. If the subset contains N vectors, the corresponding Markovian $N \times N$ matrix has N^2 entries. Consequently, instead of $M \cdot S + 2^M + 2^{2M}$, only $M \cdot S + N + N^2$ parameters have to be estimated. The heuristics and the estimators for identifying a set of significant binary vectors constitute the contribution of this paper in the literature on modeling of depending credit-rating migrations and business cycles.

The paper is structured as follows. Maximum likelihood estimators for static and dynamic models are given in Section 2. A discussion of computation and conceptual problems associated with estimating of dynamic models leads to a heuristics meant for reducing of the number of unknowns. Section 3 describes the input data and justifies the choice of M and industries for testing the models. The estimated parameters are presented and discussed in Section 4. Comparing quantitatively the heuristic solutions with the solutions for the static and the dynamic models, Section 5 concludes.

2 Why do we need a heuristics?

Let there be $S \geq 1$ industry sectors and $M \geq 1$ non-default credit classes. Consider likelihood functions and estimators corresponding to the known coupling schemes. The models are numbered by $i = 1, 2, 3$, according to the chronological order of the corresponding publications, beginning with the coupling scheme suggested in Kaniovski and Pflug (2007). Since tendency variables do not evolve in time, we call such models static. The likelihood functions L_i are:

$$L_i(\vec{\pi}, Q) = \Delta \times l_i(\vec{\pi}, Q),$$

where

$$l_1(\vec{\pi}, Q) = \prod_{t=1}^{T} \sum_{i=1}^{2^M} \pi_i \prod_{m_1=1}^{M} g_1(t, \vec{\chi}^{(i)}, m_1, Q),$$

$$l_2(\vec{\pi}, Q) = \prod_{t=1}^{T} \sum_{i=1}^{2^M} \pi_i \prod_{s=1}^{S} \prod_{m_1=1}^{M} \prod_{m_2=1}^{M+1} g_2(s, \vec{\chi}^{(i)}, m_1, m_2, Q)^{I^t(s,m_1,m_2)},$$

$$l_3(\vec{\pi}, Q) = \prod_{t=1}^{T} \sum_{i=1}^{2^M} \pi_i \prod_{s=1}^{S} \prod_{m_1=1}^{M} g_3(t, s, \vec{\chi}^{(i)}, m_1, Q),$$

$$g_1(t, \vec{\chi}, m_1, Q) = \sum_{i=1}^{M+1} P_{m_1,i}(\chi_{m_1}) \prod_{s=1}^{S} (q_{m_1,s} + \frac{1 - q_{m_1,s}}{P_{m_1,i}})^{I^t(s,m_1,i)} \prod_{m_2=1, m_2 \neq i}^{M+1} q_{m_1,s}^{I^t(s,m_1,m_2)},$$

$$g_2(s, \vec{\chi}, m_1, m_2, Q) = \begin{cases} \frac{1-q_{m_1,s}(1-P_{m_1})}{P_{m_1}}, & \text{if } m_1 \geq m_2,\ \chi_{m_1} = 1, \\ \frac{1-q_{m_1,s}P_{m_1}}{1-P_{m_1}}, & \text{if } m_1 < m_2,\ \chi_{m_1} = 0, \\ q_{m_1,s}, & \text{otherwise}, \end{cases}$$

$$g_3(t, s, \vec{\chi}, m_1, Q) = \sum_{m_2=1}^{M+1} P_{m_1,m_2}(\chi_{m_1}) (q_{m_1,s} + \frac{1 - q_{m_1,s}}{P_{m_1,m_2}})^{I^t(s,m_1,m_2)} \prod_{j=1,\, j \neq m_2}^{M+1} q_{m_1,s}^{I^t(s,m_1,j)},$$

$$\Delta = \prod_{t=1}^{T}\prod_{s=1}^{S}\prod_{m_1=1}^{M}\prod_{m_2=1}^{M+1} P_{m_1,m_2}^{I^t(s,m_1,m_2)}.$$

Containing probabilities π_l of macroeconomic scenarios $\vec{\chi}^{(l)}$ as coordinates, the vector $\vec{\pi}$ defines a probability distribution on the set of all binary M-vectors. Q is an $M \times S$ matrix with elements $q_{l,s}$. Conceptually, $q_{l,s}$ and $1-q_{l,s}$ are the weights that determine the impact of the idiosyncratic and the common component in a migration of a debtor belonging to a credit class l and an industry s. See, for details, Kaniovski and Pflug (2007), Wozabal and Hochreiter (2012) and Boreiko et. al (2016). P denotes a Markovian $(M+1)\times(M+1)$ historical migration matrix. Its entry $P_{l,j}$ equals the probability that a debtor belonging to credit class l at time t will move to credit rating j at time $t+1$. Because defaulted firms receive the index $M+1$, $M+1$ is an absorbing state of this time-homogeneous Markov chain. That is, $P_{M+1,M+1} = 1$ and $P_{M+1,j} = 0,\ j = 1,2,\ldots,M$. Matrix P is known. Typically, only entries $P_{l,j},\ l = 1,2,\ldots,M,\ j = 1,2,\ldots,M+1$, are quoted.

The coordinate χ_l of a binary vector $\vec{\chi}$ affects the evolution of debtors from credit class l modifying their migrations probabilities. The corresponding conditional distribution $P_{l,j}(\cdot)$ is defined by the following formulas:

$$P_{l,j}(1) = \begin{cases} P_{l,j}/P_l & \text{if } j \le l, \\ 0 & \text{if } j > l; \end{cases} \quad \text{and} \quad P_{l,j}(0) = \begin{cases} P_{l,j}/(1-P_l) & \text{if } j > l, \\ 0 & \text{if } j \le l. \end{cases}$$

Here, $P_l = P_{l,l} + P_{l,l-1} + \ldots + P_{l,1},\ l = 1,2,\ldots,M$. $I^t(s,m_1,m_2)$ denotes the number of debtors in industry sector s that moved from credit class m_1 to credit class m_2 in period t. T stands for the total number of periods of time, measured typically in quarters of a year or in years.

In sum, entries of Q and coordinates of $\vec{\pi}$ are the unknown parameters. Because the term Δ does not contain unknowns, it can be ignored. Estimates for $\vec{\pi}$ and Q are obtained by maximizing $\ln l_i(\vec{\pi}, Q)$ subject to linear inequality constraints, $\pi_l \in [0,1]$ and $q_{l,s} \in [0,1]$, as well as linear equality constrains,

$$\sum_{j=1}^{2^M} \pi_j = 1 \quad \text{and} \quad \sum_{j=1}^{2^M} \pi_j \mathbb{I}_{\{\chi_l^{(j)}=1\}} = P_l, \quad l = 1,2,\ldots,M.$$

Here $\mathbb{I}_{\{A\}}$ denotes the indicator function of a statement A,

$$\mathbb{I}_{\{A\}} = \begin{cases} 1 & \text{if } A \text{ holds true}, \\ 0 & \text{if } A \text{ is false}. \end{cases}$$

In the reality, macroeconomic conditions continuously change: expansions follow recessions and vise versa. There is a stream of literature in financial modeling that accounts for the effect of changing macroeconomic conditions on credit rating migrations. The coupling schemes with a Markovian dynamics presented next, generalize the class of credit-risk models entailing a regime-switching matrix. See, among others, Bangia et al. (2002), Frydman and Schuermann (2008), Fei et al. (2012).

Let $\vec{\Pi}^t$ be the a stochastic binary M-vector. We call it a tendency vector and its coordinates are referred to as tendency variables. Then $\vec{\chi}^{(i)}$, $i = 1, 2, \ldots, 2^M$, are realizations of a tendency vector. To define a Markovian dynamics of tendency vectors, a $2^M \times 2^M$ Markovian matrix $\mathcal{P}$ and a distribution of $\vec{\Pi}^1$ are necessary. The entry $p_{i,j}$ of $\mathcal{P}$ equals the probability that the macroeconomic outcome associated with $\vec{\chi}^{(j)}$ follows the macroeconomic outcome associated with $\vec{\chi}^{(i)}$:

$$p_{i,j} = \mathbb{P}\{\vec{\Pi}^{t+1} = \vec{\chi}^{(j)} \mid \vec{\Pi}^t = \vec{\chi}^{(i)}\}.$$

The distribution of $\vec{\Pi}^t$, $t \geq 2$, given that the distribution of $\vec{\Pi}^1$ is $\vec{\pi}$, will be $\vec{\pi}\mathcal{P}^{t-1}$. Here, $\mathcal{P}^{t-1}$ stands for the $(t-1)$-th power of $\mathcal{P}$. In particular,

$$\mathbb{P}\{\vec{\Pi}^t = \vec{\chi}^{(j)}\} = [\vec{\pi}\mathcal{P}^{t-1}]_j,$$

where $[\vec{\pi}\mathcal{P}^{t-1}]_j$ denotes the j-th coordinate of $\vec{\pi}\mathcal{P}^{t-1}$. Note that every static model is a particular case of the corresponding dynamic model. It corresponds to $\mathcal{P}$ equal to the $2^M \times 2^M$ identity matrix I_{2^M}. This is a trivial dynamics in the sense that every macroeconomic outcome follows itself with certainty.

Consider the likelihood functions of dynamic models:

$$L_i(\vec{\pi}, Q, \mathcal{P}) = \Delta \times l_i(\vec{\pi}, Q, \mathcal{P}),$$

$$l_1(\vec{\pi}, Q, \mathcal{P}) = \prod_{t=1}^{T} \sum_{i=1}^{2^M} [\vec{\pi}\mathcal{P}^{t-1}]_i \prod_{m_1=1}^{M} g_1(t, \vec{\chi}^{(i)}, m_1, Q),$$

$$l_2(\vec{\pi}, Q, \mathcal{P}) = \prod_{t=1}^{T} \sum_{i=1}^{2^M} [\vec{\pi}\mathcal{P}^{t-1}]_i \prod_{s=1}^{S} \prod_{m_1=1}^{M} \prod_{m_2=1}^{M+1} g_2(s, \vec{\chi}^{(i)}, m_1, m_2, Q)^{I^t(s,m_1,m_2)},$$

$$l_3(\vec{\pi}, Q, \mathcal{P}) = \prod_{t=1}^{T} \sum_{i=1}^{2^M} [\vec{\pi}\mathcal{P}^{t-1}]_i \prod_{s=1}^{S} \prod_{m_1=1}^{M} g_3(t, s, \vec{\chi}^{(i)}, m_1, Q).$$

$\vec{\pi}$ estimated here is the distribution of $\vec{\Pi}^1$. As compared with a static model, there are additionally 2^{2M} unknown parameters – entries of $\mathcal{P}$.

Estimates for $\vec{\pi}$, Q and $\mathcal{P}$ are obtained by maximizing $\ln l_i(\vec{\pi}, Q, \mathcal{P})$. There are linear equality constraints,

$$\sum_{j=1}^{2^M} \pi_j = 1 \quad \text{and} \quad \sum_{j=1}^{2^M} p_{l,j} = 1, \quad l = 1, 2, \ldots, 2^M,$$

and non-linear (with respect to $\mathcal{P}$) equality constraints,

$$\sum_{j=1}^{2^M} [\vec{\pi}\mathcal{P}^{t-1}]_j \mathbb{I}_{\{\chi_l^{(j)}=1\}} = P_l, \quad l = 1, 2, \ldots, M, \quad t = 1, 2, \ldots, T.$$

In addition to the linear inequality constraints, $\pi_l \in [0,1]$, $q_{l,s} \in [0,1]$ and $p_{l,j} \in [0,1]$, there are non-linear inequality constraints:

$$[\vec{\pi}\mathcal{P}^{t-1}]_j \leq D\pi_j, \quad j = 1, 2, \ldots, 2^M, \quad t = 2, 3, \ldots, T.$$

These non-linear inequality constraints prevent transitions to unattainable realizations of tendency vectors. We say that a binary M-vector $\vec{\chi}^{(j)}$ is unattainable, if the corresponding π_j equals 0. D is a sufficiently large positive constant.

Each of the likelihood functions corresponds to a mixture of multinomial distributions. According to Allman et al. (2009), estimating parameters of a multinomial distribution is a difficult problem: there are multiple solutions. In our case, the huge number of unknowns is an additional complication. In fact, the number of industry sectors S typically does not exceed 12. The basic S&P's classification distinguishes $M = 9$ non-default credit ratings: AAA, AA, A, BBB, BB, B, CCC, CC and C. With a reduction, merging CCC, CC and C in a single credit class C, $M = 7$ credit ratings are often considered in the literature. In this case, there are $7 \cdot S + 2^7 = 7 \cdot S + 128$ and $7 \cdot S + 2^7 + 2^{2 \cdot 7} = 7 \cdot S + 16512$ parameters to estimate for a static and, respectively, a dynamic model. A desktop computer would require a couple of minutes in order to evaluate parameters of a static model, while estimating parameters of a dynamic model is a hard task – a single iteration of an optimization algorithm could require many hours. A less apparent problem is the number of available transition counts. To obtain reliable estimates, it has to substantially exceed the number of unknown parameters.

There is a conceptual argument demonstrating that only a few macroeconomic outcomes or, equivalently, binary vectors are relevant for risk analysis. In fact, every practical problem involves a threshold probability ϵ such that random events less likely than ϵ can be ignored. Then, given a threshold ϵ, at most ϵ^{-1} binary vectors should be considered. Assuming equally probable elementary outcomes, this is a very conservative estimate. Considering a set of elementary outcomes such that one of them occurs with probability $1-\epsilon$ would be a better approach. For every ϵ, each of such sets is contained in the support of $\vec{\pi}$. Consequently, only binary vectors from the support of $\vec{\pi}$ are relevant for risk analysis

According to the heuristics described next, the state space of a dynamic model is limited to the support of the corresponding static model. If the support contains N sample points, an $N \times N$ Markovian matrix governing the macroeconomic dynamics has to be estimated. This heuristic solution is always better than the corresponding static solution, but, typically, it is not optimal. In particular, it cannot be optimal if the support of a static model is a proper subset of the support of the corresponding dynamic model. In any case, the heuristics is less demanding than the exact model in what concerns the computational resources. An attractive feature of this simplification is that a heuristic solution can be quantitatively compared with its static counterpart.

Let $(Q^{(i)}, \vec{\pi}^{(i)})$ be parameters estimated for the static model i. Consider the support $TV^{(i)} = \{\vec{\chi}^{(j_1)}, \vec{\chi}^{(j_2)}, \ldots, \vec{\chi}^{(j_{N_i})}\}$ of the distribution $\vec{\pi}^{(i)}$. Introduce likelihood functions $L_i^0(\vec{d}, \bar{Q}, \mathcal{R})$ that are nested exclusively on binary vectors from $TV^{(i)}$:

$$L_i^0(\vec{d}, \bar{Q}, \mathcal{R}) = \Delta \times l_i^0(\vec{d}, \bar{Q}, \mathcal{R}),$$

$$l_1^0(\vec{d},\bar{Q},\mathcal{R}) = \prod_{t=1}^{T}\sum_{k=1}^{N_1}[\vec{d}\mathcal{R}^{t-1}]_k \prod_{m_1=1}^{M} g_1(t,\vec{\chi}^{(j_k)},m_1,\bar{Q}),$$

$$l_2^0(\vec{d},\bar{Q},\mathcal{R}) = \prod_{t=1}^{T}\sum_{k=1}^{N_2}[\vec{d}\mathcal{R}^{t-1}]_k \prod_{s=1}^{S}\prod_{m_1=1}^{M}\prod_{m_2=1}^{M+1} g_2(s,\vec{\chi}^{(j_k)},m_1,m_2,\bar{Q})^{I^t(s,m_1,m_2)},$$

$$l_3^0(\vec{d},\bar{Q},\mathcal{R}) = \prod_{t=1}^{T}\sum_{k=1}^{N_3}[\vec{d}\mathcal{R}^{t-1}]_k \prod_{s=1}^{S}\prod_{m_1=1}^{M} g_3(t,s,\vec{\chi}^{(j_k)},m_1,\bar{Q}).$$

Here, $\vec{d}$ is a vector with N_i coordinates which defines a distribution on $TV^{(i)}$: d_k is the probability assigned to $\vec{\chi}^{(j_k)}$. $\bar{Q}$ denotes an $M \times S$ matrix with elements $\bar{q}_{l,s}$. $\mathcal{R}$ is an $N_i \times N_i$ Markovian matrix whose entry $r_{I,J}$ equals the probability that the state of economy characterized by $\vec{\chi}^{(j_I)}$ at time t will be followed by the state corresponding to $\vec{\chi}^{(j_J)}$ at $t+1$. (To avoid bulky notations, the index i is omitted in $\vec{d}$, $\bar{Q}$ and $\mathcal{R}$.)

Estimates $\vec{d}^{(i)}$, $\bar{Q}^{(i)}$ and $\mathcal{R}^{(i)}$ are obtained by maximizing $\ln l_i^0(\vec{d},\bar{Q},\mathcal{R})$ subject to linear equality constraints,

$$\sum_{k=1}^{N_i} d_k = 1 \quad \text{and} \quad \sum_{I=1}^{N_i} r_{I,J} = 1, \quad J = 1,2,\ldots,N_i,$$

and non-linear equality constraints,

$$\sum_{k=1}^{N_i}[\vec{d}\mathcal{R}^{t-1}]_j \mathbb{I}_{\{\chi_j^{(i_k)}=1\}} = P_j, \quad j = 1,2,\ldots,M, \quad t = 1,2,\ldots,T.$$

There are also linear inequality constraints, $d_j \in [0,1]$, $\bar{q}_{i,s} \in [0,1]$ and $r_{I,J} \in [0,1]$, as well as non-linear inequality constraints,

$$[\vec{d}\mathcal{R}^{t-1}]_j \le D d_j, \quad j = 1,2,\ldots,N_i, \quad t = 2,3,\ldots,T.$$

A solution obtained by this heuristics will not be, in general, an optimal solution for the corresponding dynamic setting. However, since N_i^2 is much smaller than 2^{2M}, the estimation problem is simpler.

Interpreting coordinates of a tendency vector as a binary representation of an integer, let us number the binary strings in a descending order of the corresponding integers assigning 1 to the binary vector with al coordinates equal to 1. In particular, if $M = 2$, four elements of $\{0,1\}^2$ appear in the following order:

$$\vec{\chi}^{(1)} = (1,\ 1), \quad \vec{\chi}^{(2)} = (1,\ 0), \quad \vec{\chi}^{(3)} = (0,\ 1), \quad \vec{\chi}^{(4)} = (0,\ 0).$$

3 Input data

A S&P's dataset containing credit-ratings migrations from 1991 through 2012 is used. That is, time instant $1(T)$ corresponds to 1991 (2012). There are 94584 counts altogether.

The debtors belong to the OECD (Organization for Economic Co-operation and Development) countries. Two combinations of M and S are considered: $M = 7$ with $S = 1$ and $M = 2$ with $S = 12$.

If $M = 7$, the S&P's credit classes AAA, AA, A, BBB, BB, B and C are numbered by $1, 2, \ldots, 7$. Considering two non-default credit classes, index 1(2) is assigned to investment-grade (non-investment-grade) debtors. The investment-grade debtors are characterized by the S&P's ratings from AAA to BBB. The non-investment-grade debtors are those who received a riskier rating, BB, B or C.

If $S = 1$ the debtors are not classified according to their industry. If $S = 12$, the following industries are considered: 1 – aero, auto, capital goods, metal; 2 – consumer, service; 3 – energy, natural resources; 4 – financial institutions; 5 – forest and building products, homebuilders; 6 – health care, chemicals; 7 – high technology, computers, office equipment; 8 – insurance, real estate; 9 – leisure time, media; 10 – telecommunications; 11 – transportation; 12 – utilities.

Given next matrices P were estimated as the time averages of the respective transition counts:

$$\begin{pmatrix} 0.8952 & 0.0981 & 0.0047 & 0.0008 & 0.0000 & 0.0000 & 0.0000 & 0.0011 \\ 0.0064 & 0.8983 & 0.0895 & 0.0045 & 0.0002 & 0.0007 & 0.0002 & 0.0001 \\ 0.0010 & 0.0362 & 0.9022 & 0.0566 & 0.0020 & 0.0007 & 0.0004 & 0.0010 \\ 0.0013 & 0.0049 & 0.0567 & 0.8806 & 0.0476 & 0.0065 & 0.0010 & 0.0015 \\ 0.0006 & 0.0033 & 0.0100 & 0.1115 & 0.7862 & 0.0754 & 0.0055 & 0.0074 \\ 0.0008 & 0.0013 & 0.0044 & 0.0112 & 0.0949 & 0.8015 & 0.0520 & 0.0340 \\ 0.0015 & 0.0000 & 0.0015 & 0.0030 & 0.0213 & 0.1393 & 0.5693 & 0.2641 \end{pmatrix},$$

$$\begin{pmatrix} 0.9776 & 0.0214 & 0.0010 \\ 0.0746 & 0.8935 & 0.0319 \end{pmatrix}.$$

4 Estimates

According to Allman et al. (2009), multiple solutions is a common complication while estimating parameters of multinomial mixtures. In the case of a coupling scheme, applying the maximum likelihood method, two remedies for avoiding local minima have been reported in the literature: the repeated use of different algorithms from an optimization package in Boreiko et. al (2016) and the particle swarm algorithm, a heuristic global optimization method, in Wozabal and Hochreiter (2012). We used two MATLAB solvers, the Interior Point (IP) method and the Sequential Quadratic Programming (SQP) algorithm, to estimate the parameters. For every model, both methods found the same solution. Since a variety of initial approximations have been tried, it can be an indication that a global maximum of the likelihood function was found. This positive experience is consistent with what other researchers report about successful numerical implementations of the maximum likelihood estimators for mixtures of multinomial, in particular binomial, distributions. See, for example, Carreira-Perpiñán and Renals (2000). In general, the success of applying the IP algorithm to a dynamic estimator depends stronger, than when the SQP method is used, on the choice of an initial approximation.

Note that the models 1 and 3 coincide if $S = 1$. First let us present estimates for the static models.

$S = 1$ and $M = 7$. Matrices $Q^{(1)}$ and $Q^{(2)}$:

$$(0.8359, 0.9106, 0.7984, 0.9061, 0.8404, 0.9010, 0.7708),$$

$$(0.8322, 0.8647, 0.9583, 0.9545, 0.9545, 0.9067, 0.7279).$$

Supports of distributions $\vec{\pi}^{(1)}$ and $\vec{\pi}^{(2)}$ contain 11 and 13 sample points. They are characterized by the following relations:

$$\pi_1^{(1)} = 0.5723,\ \pi_2^{(1)} = 0.1130,\ \pi_8^{(1)} = 0.0277,\ \pi_9^{(1)} = 0.0566,\ \pi_{22}^{(1)} = 0.0502,$$

$$\pi_{24}^{(1)} = 0.0104,\ \pi_{33}^{(1)} = 0.0151,\ \pi_{34}^{(1)} = 0.0324,\ \pi_{35}^{(1)} = 0.0176,\ \pi_{65}^{(1)} = 0.0745,$$

$$\pi_{100}^{(1)} = 0.0304;\ \ \pi_1^{(2)} = 0.5343,\ \pi_2^{(2)} = 0.1756,\ \pi_6^{(2)} = 0.0221,\ \pi_7^{(2)} = 0.0197,$$

$$\pi_9^{(2)} = 0.0326,\ \pi_{24}^{(2)} = 0.0220,\ \pi_{25}^{(2)} = 0.0181,\ \pi_{33}^{(2)} = 0.0363,\ \pi_{36}^{(2)} = 0.0198,$$

$$\pi_{49}^{(2)} = 0.0146,\ \pi_{65}^{(2)} = 0.0803,\ \pi_{104}^{(2)} = 0.0187,\ \pi_{128}^{(2)} = 0.0058.$$

In fact, $\pi_{36}^{(2)} = 0.0198$, by our numbering of binary vectors, implies that according to model 2 probability 0.0198 is assigned to the binary vector $\vec{\chi}^{(36)} = (1, 0, 1, 1, 1, 0, 0)$. The corresponding macroeconomic outcome is favorable for debtors belonging to the credit classes AAA, A, BBB and BB, while it is adverse for debtors rated at AA, B and C.

$S = 12$ and $M = 2$. Matrices $Q^{(i)}$:

$$\begin{pmatrix} 1.0000 & 1.0000 & 0.5531 & 0.3337 & 1.0000 & 0.9793 & 1.0000 & 1.0000 & 1.0000 & 0.9307 & 0.7852 & 0.6026 \\ 0.9833 & 0.9767 & 0.9924 & 0.9364 & 0.9567 & 0.9492 & 0.9433 & 0.8107 & 0.9677 & 0.9124 & 0.9794 & 0.9741 \end{pmatrix},$$

$$\begin{pmatrix} 1.0000 & 1.0000 & 0.5531 & 0.3337 & 1.0000 & 0.9793 & 1.0000 & 1.0000 & 1.0000 & 0.9307 & 0.7852 & 0.6026 \\ 0.9318 & 0.9643 & 0.9760 & 0.9444 & 0.9283 & 0.9833 & 0.8862 & 0.9840 & 0.9613 & 0.8108 & 0.9591 & 0.9520 \end{pmatrix},$$

$$\begin{pmatrix} 1.0000 & 1.0000 & 0.5531 & 0.3337 & 1.0000 & 0.9793 & 1.0000 & 1.0000 & 1.0000 & 0.9307 & 0.7852 & 0.6026 \\ 0.6520 & 0.5680 & 0.4241 & 1.0000 & 0.6659 & 0.4951 & 0.7583 & 0.8083 & 0.4795 & 1.0000 & 0.6247 & 0.7857 \end{pmatrix}.$$

Distributions $\vec{\pi}^{(1)}$ and $\vec{\pi}^{(2)}$ are equal. The support of $\vec{\pi}^{(1)}$ as well as the support of $\vec{\pi}^{(3)}$ contains 3 sample points. The supports do not coincide, since

$$\pi_1^{(1)} = 0.9457,\ \pi_2^{(1)} = 0.0319,\ \pi_3^{(1)} = 0.0224,$$

while

$$\pi_1^{(3)} = 0.9681,\ \pi_2^{(3)} = 0.0095,\ \pi_4^{(3)} = 0.0224.$$

In fact, the support of $\vec{\pi}^{(1)}$ consists of (1, 1), (1, 0) and (0, 1), while the supports of $\vec{\pi}^{(3)}$ contains (0, 0) instead of (0, 1).

Testing dynamic models, parameters of the exact models and of the heuristics were estimated in the case of $S = 12$ and $M = 2$. The penalty parameter $D = 200$ was used in all models. For $S = 1$ and $M = 7$, only parameters corresponding to the heuristics were estimated. The value of D used in the calculations was 30000.

$\underline{S = 1 \text{ and } M = 7}$. Matrices $\bar{Q}^{(1)}$ and $\bar{Q}^{(2)}$ equal to

$$(0.8359, 0.9109, 0.7984, 0.9072, 0.8404, 0.9008, 0.7737),$$

and

$$(0.8352, 0.8622, 0.9583, 0.9546, 0.9546, 0.9068, 0.7337).$$

These values resemble closely their counterparts for the static models.

Probabilities $d_i^{(1)}$ and $d_i^{(2)}$ are:

$$d_1^{(1)} = 0.5237,\ d_2^{(1)} = 0.1400,\ d_3^{(1)} = 0.0277,\ d_4^{(1)} = 0.0566,\ d_5^{(1)} = 0.0344,$$
$$d_6^{(1)} = 0.0262,\ d_7^{(1)} = 0.0362,\ d_8^{(1)} = 0.0270,\ d_9^{(1)} = 0.0234,\ d_{10}^{(1)} = 0.0961,$$
$$d_{11}^{(1)} = 0.0087;\ d_1^{(2)} = 0.4828,\ d_2^{(2)} = 0.2227,\ d_3^{(2)} = 0.0092,\ d_4^{(2)} = 0.0539,$$
$$d_5^{(2)} = 0.0082,\ d_6^{(2)} = 0.0030,\ d_7^{(2)} = 0.0422,\ d_8^{(2)} = 0.0569,\ d_9^{(2)} = 0.0070,$$
$$d_{10}^{(2)} = 0.0092,\ d_{11}^{(2)} = 0.0826,\ d_{12}^{(2)} = 0.0161,\ d_{13}^{(2)} = 0.0061.$$

These distributions are quite similar to their counterparts of the static models. (Recall that supports of $\vec{d}^{(i)}$ and $\vec{\pi}^{(i)}$ coincide. In particular, $d_3^{(1)}$ is assigned to the binary vector (1, 1, 1, 1, 0, 0, 0) listed third in the support of $\vec{\pi}^{(1)}$. The corresponding probability for the static model 1 equals $\pi_8^{(1)} = 0.0277$.) Markovian matrices $\mathcal{R}^{(1)}$ and $\mathcal{R}^{(2)}$,

$$\begin{pmatrix}
0.6826 & 0.1212 & 0.0209 & 0.0376 & 0.0052 & 0.0191 & 0.0293 & 0.0131 & 0.0121 & 0.0514 & 0.0076 \\
0.5235 & 0.3640 & 0.0078 & 0.0149 & 0.0145 & 0.0142 & 0.0126 & 0.0241 & 0.0106 & 0.0115 & 0.0022 \\
0.1216 & 0.0373 & 0.2247 & 0.0435 & 0.0445 & 0.0554 & 0.1460 & 0.0342 & 0.1261 & 0.1567 & 0.0101 \\
0.3342 & 0.1684 & 0.0332 & 0.1812 & 0.0074 & 0.0879 & 0.0552 & 0.0473 & 0.0299 & 0.0476 & 0.0077 \\
0.0780 & 0.0066 & 0.0113 & 0.0177 & 0.6882 & 0.0334 & 0.0609 & 0.0240 & 0.0072 & 0.0586 & 0.0142 \\
0.5037 & 0.0911 & 0.0327 & 0.1915 & 0.0200 & 0.0573 & 0.0182 & 0.0408 & 0.0078 & 0.0301 & 0.0067 \\
0.2485 & 0.1634 & 0.0497 & 0.0336 & 0.0219 & 0.0317 & 0.0984 & 0.1253 & 0.0800 & 0.1333 & 0.0143 \\
0.2034 & 0.0973 & 0.0334 & 0.1309 & 0.0382 & 0.0433 & 0.0722 & 0.1344 & 0.0762 & 0.1518 & 0.0189 \\
0.3582 & 0.0739 & 0.0366 & 0.1052 & 0.0151 & 0.0631 & 0.1158 & 0.0455 & 0.1320 & 0.0387 & 0.0159 \\
0.3292 & 0.0165 & 0.0274 & 0.0931 & 0.0163 & 0.0071 & 0.0108 & 0.0083 & 0.0141 & 0.4734 & 0.0039 \\
0.0299 & 0.0657 & 0.0158 & 0.1791 & 0.0097 & 0.0666 & 0.0137 & 0.1344 & 0.0654 & 0.2721 & 0.1477
\end{pmatrix},$$

and

$$\begin{pmatrix}
0.9723 & 0.0164 & 0.0022 & 0.0035 & 0.0001 & 0.0000 & 0.0006 & 0.0027 & 0.0004 & 0.0006 & 0.0012 & 0.0001 & 0.0000 \\
0.0404 & 0.9321 & 0.0035 & 0.0010 & 0.0012 & 0.0005 & 0.0038 & 0.0016 & 0.0007 & 0.0008 & 0.0140 & 0.0003 & 0.0001 \\
0.0582 & 0.0105 & 0.6822 & 0.1671 & 0.0065 & 0.0022 & 0.0003 & 0.0151 & 0.0243 & 0.0080 & 0.0032 & 0.0215 & 0.0008 \\
0.0003 & 0.0033 & 0.0025 & 0.8804 & 0.0007 & 0.0004 & 0.0235 & 0.0113 & 0.0001 & 0.0008 & 0.0759 & 0.0007 & 0.0001 \\
0.0090 & 0.0003 & 0.0166 & 0.0145 & 0.9174 & 0.0015 & 0.0070 & 0.0059 & 0.0114 & 0.0041 & 0.0023 & 0.0012 & 0.0088 \\
0.0005 & 0.0050 & 0.0680 & 0.0076 & 0.0001 & 0.9028 & 0.0026 & 0.0042 & 0.0054 & 0.0015 & 0.0002 & 0.0022 & 0.0001 \\
0.0120 & 0.0273 & 0.0056 & 0.0310 & 0.0080 & 0.0103 & 0.8347 & 0.0297 & 0.0056 & 0.0007 & 0.0269 & 0.0039 & 0.0042 \\
0.0042 & 0.0468 & 0.0041 & 0.0068 & 0.0003 & 0.0016 & 0.0226 & 0.7897 & 0.0032 & 0.0002 & 0.1184 & 0.0013 & 0.0006 \\
0.0160 & 0.0260 & 0.0037 & 0.0056 & 0.0030 & 0.0009 & 0.0174 & 0.0303 & 0.8149 & 0.0030 & 0.0619 & 0.0154 & 0.0017 \\
0.0129 & 0.0053 & 0.0094 & 0.0036 & 0.0001 & 0.0059 & 0.1165 & 0.0006 & 0.0029 & 0.7662 & 0.0539 & 0.0026 & 0.0200 \\
0.0326 & 0.0245 & 0.0036 & 0.0027 & 0.0003 & 0.0004 & 0.0127 & 0.1079 & 0.0056 & 0.0025 & 0.8063 & 0.0005 & 0.0002 \\
0.0012 & 0.0027 & 0.0001 & 0.0270 & 0.0081 & 0.0172 & 0.0003 & 0.0002 & 0.0001 & 0.0442 & 0.0013 & 0.8702 & 0.0276 \\
0.0026 & 0.0137 & 0.0086 & 0.0009 & 0.0009 & 0.0002 & 0.0166 & 0.0052 & 0.0005 & 0.0037 & 0.0082 & 0.0014 & 0.9376
\end{pmatrix},$$

imply different macroeconomic dynamics. According to model 2, there is a strong tendency for all states to recur: the smallest diagonal entry (meaning conceptually the maintain probability) is $r_{3,3}^{(2)} = 0.6822$. Model 2 predicts this pattern only for the binary vectors (1, 1, 1, 1, 1, 1, 1) and (1, 1, 0, 1, 0, 1, 0). The first outcome is favorable for all debtors in the dataset, while the second outcome is favorable for debtors belonging to the credit classes AAA, AA, BBB, B and it is not favorable for those rated at A, BB and C.

$\underline{S = 12 \text{ and } M = 2, \text{ exact models}}$. Matrices $Q^{(i)}$:

$$\begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.9945 & 1.0000 & 1.0000 & 0.9497 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.9702 & 0.9741 & 0.9671 & 0.9487 & 0.9370 & 0.9751 & 0.9239 & 0.8337 & 0.9694 & 0.8350 & 0.9579 & 0.9601 \end{pmatrix},$$

$$\begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.9956 & 1.0000 & 1.0000 & 0.9507 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.9324 & 0.9643 & 0.9760 & 0.9445 & 0.9284 & 0.9833 & 0.8862 & 0.9840 & 0.9613 & 0.8108 & 0.9591 & 0.9521 \end{pmatrix},$$

$$\begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & 0.9956 & 1.0000 & 1.0000 & 0.9497 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.6520 & 0.5680 & 0.4241 & 1.0000 & 0.6659 & 0.4951 & 0.7583 & 0.8083 & 0.4795 & 1.0000 & 0.6247 & 0.7857 \end{pmatrix}.$$

Distributions $\vec{\pi}^{(i)}$ and Markovian matrices $\mathcal{P}^{(i)}$ are identical for the three models:

$$\pi_1^{(i)} = 0.9547,\ \pi_2^{(i)} = 0.0206,\ \pi_3^{(i)} = 0.0060,\ \pi_4^{(i)} = 0.0186;$$

$$\mathcal{P}^{(i)} = \begin{pmatrix} 0.9894 & 0.0006 & 0.0073 & 0.0026 \\ 0.0800 & 0.9098 & 0.0035 & 0.0067 \\ 0.6786 & 0.0976 & 0.1373 & 0.0866 \\ 0.2209 & 0.1613 & 0.3837 & 0.2340 \end{pmatrix}.$$

The states (1, 1) and (1, 0) tend to persist. Their maintain probabilities are 0.9894 and 0.9098.

$\underline{S = 12 \text{ and } M = 2, \text{ the heuristics}}$. Matrices $\bar{Q}^{(1)}$ and $\bar{Q}^{(2)}$ coincide with their counterparts for the exact dynamic models, while $\bar{Q}^{(3)}$ is identical to $Q^{(3)}$ estimated for the static model. Distributions $\vec{d}^{(i)}$ and Markovian matrices $\mathcal{R}^{(i)}$ are the same for all three models:

$$d_1^{(i)} = 0.9609,\ d_2^{(i)} = 0.0268,\ d_3^{(i)} = 0.0122; \quad \mathcal{R}^{(i)} = \begin{pmatrix} 0.9844 & 0.0000 & 0.0156 \\ 0.0189 & 0.9470 & 0.0341 \\ 0.6683 & 0.0225 & 0.3093 \end{pmatrix}.$$

(Note that supports of the distributions are different: $\vec{d}^{(1)}$ and $\vec{d}^{(2)}$ are nested in $(1,1),\ (1,0),\ (0,1)$, whereas $\vec{d}^{(3)}$ is nested in (1, 1), (1, 0), (0, 0).) The probabilities assigned to (1, 1) and (1, 0) are close to to their analogs estimated for the exact dynamic models. Similar to what is observed for the exact models, these states tend to persist: their maintain probabilities are 0.9844 and 0.9470.

5 Comparison of exact solutions and heuristics

The suggested heuristics does not allow, in general, to identify an optimal solution for the corresponding dynamic setting. In fact, considering $S = 12$ and $M = 2$, the support of an optimal solution $\vec{\pi}^{(i)}$ for the static setting for model i consists of three binary vectors for each of the models. Consequently, the respective heuristic dynamics are nested in three binary vectors. Their exact dynamic counterpart entails four binary vectors for each of the models. All of the estimated distributions $\vec{\pi}$ ($\vec{d}$) and the corresponding matrices $\mathcal{P}$ ($\mathcal{R}$) satisfy the equation $\vec{\pi}\mathcal{P} = \vec{\pi}$ ($\vec{d}\mathcal{R} = \vec{d}$). Consequently, all Markov chains formed by the tendency vectors are tationary and $\vec{\pi}$ ($\vec{d}$) is its steady-state distribution. In all cases, the trivial Markovian dynamics, where each realization of a tendency vector follows itself with probability one, is not optimal. This dynamics corresponds to the respective static model.

Quantitatively, the improvement due to introducing a Markovian dynamics of tendency vectors is characterized by the two tables given next. Recall that $\vec{\pi}^{(i)}$, $Q^{(i)}$ stands for an optimal solution for the static model i. By $\bar{l}_i(\bar{l}_i^0)$ we denote the optimal value of the exact (heuristic) dynamic model. The comparison is possible because $\vec{\pi}^{(i)}$, $Q^{(i)}$ combined with I_{2^M} (I_{N_i}), the identity $2^M \times 2^M$ ($N_i \times N_i$) matrix, is a feasible point for the estimator of dynamic (heuristic) model i. Tables 1 and 2 demonstrate that all dynamic solutions, exact and heuristic, are better than their static analogs. For models 1 and 2, the heuristic solution appears to be practically as good as the exact one.

Table 1. Dynamic vs. static models: likelihood ratio.

	$S = 1$ and $M = 7$		$S = 12$ and $M = 2$		
	$i = 1$	$i = 2$	$i = 1$	$i = 2$	$i = 3$
$\frac{\bar{l}_i}{l_i(\vec{\pi}^{(i)}, Q^{(i)}, I_{2^M})}$	–	–	$e^{125.7}$	$e^{108.9}$	$e^{109.7}$
$\frac{\bar{l}_i^{(0)}}{l_i^{(0)}(\vec{\pi}^{(i)}, Q^{(i)}, I_{N_i})}$	$e^{26.4}$	$e^{37.6}$	$e^{125.5}$	$e^{108.7}$	$e^{0.7}$

Table 2. Dynamic vs. static models, logarithmic scale: percentage of improvement.

	$S = 1$ and $M = 7$		$S = 12$ and $M = 2$		
	$i = 1$	$i = 2$	$i = 1$	$i = 2$	$i = 3$
$\frac{\ln \bar{l}_i - \ln l_i(\vec{\pi}^{(i)}, Q^{(i)}, I_{2^M})}{\ln l_i(\vec{\pi}^{(i)}, Q^{(i)}, I_{2^M})} 100$	–	–	15.3	33.6	11.1
$\frac{\ln \bar{l}_i^{(0)} - \ln l_i^{(0)}(\vec{\pi}^{(i)}, Q^{(i)}, I_{N_i})}{\ln l_i^{(0)}(\vec{\pi}^{(i)}, Q^{(i)}, I_{N_i})} 100$	1.8	7.8	15.2	33.5	0.1

References

Allman, E. L., Matias, C., and Rhodes, J. A. (2009). Identifiability of parameters in latent structure with many observed variables. *Annals of Statistics*, 37:3099-3132.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., and Schuermann, T. (2002). Ratings migration and the business cycle, with applications to credit portfolio stress testing. *Journal of Banking and Finance*, 26:445-474.

Boreiko, D. V., Kaniovski, Y. M., and Pflug, G. Ch. (2016). Modeling dependent credit rating transitions – a comparison of coupling schemes and empirical evidence. *Central European Journal of Operations Research,* 24:989-1007.

Carreira-Perpiñán, M., and Renals, S. (2000). Practical identifiability of finite mixtures of multivariate Bernoulli distributions. *Neural Computation*, 12:141-152.

Fei, F., Fuertes, A.-M., and Kalotychou E. (2012). Credit rating migration risk and business cycles. *Journal of Business Finance and Accounting*, 39:229-263.

Frydman, H., and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking and Finance*, 32:1062-1075.

Gupton, G. M., Finger, Ch. C., and Bhatia, M. (1997). *CreditMetrics – Technical Document*, New York: J.P. Morgan Inc.

Kaniovski, Y. M., and Pflug, G. Ch. (2007). Risk assessment for credit portfolios: A coupled Markov chain model. *Journal of Banking and Finance*, 31:2303-2323.

Korolkiewicz, M., and Elliott, R. (2008). A hidden Markov model of credit quality. *Journal of Economic Dynamics and Control*, 32:3807-3819.

McNeil, A., and Wendin, J. (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance*, 14:131-149.

Wozabal, D., and Hochreiter, R. (2012). A coupled Markov chain approach to credit risk modeling. *Journal of Economic Dynamics and Control*, 36:403-415.

Xing, H., Sun, N., and Chen, Y. (2012). Credit rating dynamics in the presence of unknown structural breaks. *Journal of Banking and Finance*, 36:78-89.