STOCHASTIC MIGRATION MODELS WITH APPLICATION TO CORPORATE RISK

P. Gagliardini[*], C. Gouriéroux[†]

June 4, 2004

[*]Università della Svizzera Italiana, and CREST.
[†]CREST, CEPREMAP, and University of Toronto.

We thank P. Balestra, G. Laroque and T. Schuerman for helpful comments.

Stochastic Migration Models with Application to Corporate Risk

Abstract

We consider a set of Markovian processes with stochastic transition matrices. This specification extends the standard stochastic intensity model introduced by Cox in the two state case. Such a model is appropriate for the joint analysis of rating histories of several corporates, including the link between upgrades or downgrades, that is the so-called migration correlation. Different specifications of the micro- and macro-components of the model are introduced and discussed. As an illustration the ordered Probit model with unobservable dynamic factor is estimated from French data on corporate risk.

Keywords: Migration, Rating, Migration Correlation, Credit Risk, Stochastic Intensity, Autoregressive Gamma Process, Jacobi Process, Ordered Qualitative Model, Kalman Filter, Panel Data.

JEL number: C23, C35, G11.

Modèles de migration stochastique avec application au risque de crédit des entreprises

Résumé

Nous considérons un ensemble de processus de Markov avec des matrices de transition stochastiques. Cette spécification étend le modèle à intensité stochastique introduit par Cox dans le cas de deux états. Un tel modèle est approprié pour analyser les historiques joints de notation (rating) de diverses entreprises, incluant les liens entre amélioration ou dégradation des notes, c'est à dire la corrélation de migration. Diverses spécifications des composantes micro- et macroéconomiques du modèle sont introduites et discutées. A titre d'illustration nous estimons un modèle probit ordonné à facteur non observable à partir de données françaises sur le risque de crédit des entreprises.

Mots clés: migration, rating, corrélation de migration, risque de crédit, intensité stochastique, processus autorégressif gamma, processus de Jacobi, modèle qualitatif ordonné, filtre de Kalman, données de panel.

Classification JEL: C23, C35, G11.

1 Introduction

This paper considers a set of Markovian processes with stochastic transition matrices. This specification extends the standard stochastic intensity model[1] introduced by Cox in the two state case, and used in the financial literature for analyzing credit risk and default correlation[2]. Such a specification is especially appropriate for the joint analysis of rating histories of several corporates, including the link between rating upgrades or downgrades, the so-called migration correlation.

In Section 2 we define the basic model for stochastic migration. In this model the individual qualitative rating histories are independent heterogeneous Markov processes with identical time varying stochastic transition matrices. The underlying process of transition matrices acts as a multivariate systematic factor which affects all individual histories and creates the correlation between the histories. We discuss the predictive properties of the model according to the prediction horizon and to the available information set. Different specifications for the dynamics of the stochastic transition matrices are discussed in Section 3. They include the case of i.i.d. transition matrices, ordered probit model with serial correlation or reduced form modelling via the Jacobi process. Section 4 focuses on the definition of migration correlation, which is a corner stone in credit risk analysis. Indeed this notion has not been precisely defined in the academic or applied literature; whereas some estimators have been proposed and are regularly reported by the rating agencies, they are computed "without relying on a specific model driving transitions" [de Servigny, Renault (2002)]. This can explain the following remark in the seminal paper by Lucas (1995a) p82: "These historical statistics describe only observed phenomena, not the true underlying correlation relationship". Precise definitions of migration correlations are provided in the framework of stochastic migration models. We explain how they are related to the dynamics of the underlying transition matrices and discuss the restrictions on the term structure of migration correlation coming from the dynamic assumptions. Section 5 is concerned with statistical inference. Since the model with stochastic transition introduced in this paper is a nonlinear factor model for panel data, standard simulation based estimation methods can be used. This explains why the section will focus on some special features

[1] Also known as the doubly stochastic intensity model.

[2] See e.g. Lando (1998), Duffie, Singleton (1999), Duffie, Lando (2001), Gouriéroux, Monfort, Polimenis (2003).

such as i) the consistency of the ML estimator when either the cross-sectional or the time dimension tends to infinity, ii) the problem of default absorbing barrier, iii) the implementation of an approximate Kalman filter for large portfolios. Finally Section 6 presents an application to the migration data regularly reported by the French central bank. We display the estimated migration correlations and compare the estimated default correlations with the values currently suggested by the regulator. Then we discuss the number of dynamic factors and their relationship with the GNP increments (business cycle) in terms of causality analysis. Finally we consider an ordered probit transition model with unobservable factor and perform its estimation by an approximated Kalman filter. To our knowledge this is one of the first estimation of such a model [see Feng et alii (2003) for a similar analysis on Standard and Poor's data]. Section 7 concludes.

2 The basic factor model

The aim of this section is to introduce a factor model for a joint analysis of a large number of qualitative individual histories $(Y_{i,t})$, $i = 1, ..., n$ with the same known finite state space $\{1, ..., K\}$. For the credit risk applications the individual can be a firm, the states $k = 1, ..., K$ correspond to the admissible grades such as AAA, AA, ..., D, and a given process $(Y_{i,t})$ to a sequence of individual ratings over time. The model is defined in Section 2.1 and the homogeneity assumption discussed in Section 2.2 for credit risk analysis. Section 2.3 is concerned by the predictive properties of the stochastic migration model. We distinguish carefully probabilistic predictions and analysis by scenarios, and discuss the effect of the available information set.

2.1 Definition

The joint dynamics of individual histories is defined as follows.

Definition 1: The individual histories satisfy a Markov stochastic transition (MST), or stochastic migration model if:

i) the processes $(Y_{i,t})$, $i = 1, ..., n$, are independent Markov chains, with identical transition matrices Π_t, when the sequence of transitions (Π_t) is given;

ii) the process of transition matrices (Π_t) is a stochastic Markov process.

In practice the transition matrices are generally written as functions of a small number of factors $\Pi_t = \Pi(Z_t)$, say, satisfying a Markov process. Under a stochastic migration model the whole dependence between individual histories is driven by the common factor Z_t (or Π_t).

The stochastic migration model is a convenient specification to get joint histories featuring cross-individual dependence. Indeed the model could have been defined directly for the joint process of (rating) histories (Y_t), where $Y_t = (Y_{1,t}, ..., Y_{n,t})'$. However, since the number of admissible states for Y_t is K^n, and the conditional distribution of Y_{t+1} may depend potentially on the whole past history $\underline{Y_t}$ [3], such a direct approach would require the specification of a very large number of transition probabilities. Even under the assumption that the joint process of (rating) histories (Y_t) is Markov[4], the associated joint transition matrix includes $K^n(K^n - 1)$ independent transition probabilities. For instance in the standard application to corporate rating, the number of admissible grades AAA, AA, ..., D is about $K \simeq 10$ [5], and the number of rated corporates is between $n = 10000$ and $n = 100000$ (about 10000 for the S & P data base concerning the rather large international corporates [see Brady, Bos (2002)], and 180000 in the data base of the French Central Bank which follows all French corporates [see Foulcher et alii (2003)]); thus we will immediately encounter the curse of dimensionality. The MST specification is introduced to constrain the transitions and to diminish the number of parameters to be estimated. They will include the parameters characterizing the dependence between the transition probabilities Π_t and the factors Z_t, plus the parameters defining the factor dynamics.

The dynamics of the chains $(Y_{i,t})$ can be analyzed in alternative ways according to the available information.

i) If the past, current and future values of the underlying factors are observed, the processes $(Y_{i,t})$ are independent Markov chains. They are non stationary, since the transition matrices differ in time.

ii) If the underlying factors are not observed, it is necessary to integrate out the factors (Z_t) [or the transition matrices (Π_t)]. Let us discuss the joint

[3] $\underline{Y_t}$ denotes the current and lagged ratings: $\underline{Y_t} = (Y_t, Y_{t-1}, Y_{t-2}, ...)$.

[4] See Section 3.1 for the discussion of this special case.

[5] For the regulation by the Basle Committee the number of grades has to be between 8 and 10.

distribution of Y_{t+1} given the lagged ratings $\underline{Y_t}$ only. Its transition matrix is characterised by:

$$\begin{aligned} & P\left(Y_{1,t+1}=k_1^*,...,Y_{n,t+1}=k_n^* \mid \underline{Y_t}\right) \\ = \; & E\left[P\left(Y_{1,t+1}=k_1^*,...,Y_{n,t+1}=k_n^* \mid \underline{Y_t},(\Pi_t)\right) \mid \underline{Y_t}\right] \\ = \; & E\left[\pi_{k_1k_1^*,t+1}...\pi_{k_nk_n^*,t+1} \mid \underline{Y_t}\right], \quad \text{where } Y_{1,t}=k_1,...,Y_{n,t}=k_n. \end{aligned}$$

We deduce the property below:

Proposition 1: For a Markov stochastic transition model, the distribution of the joint process (Y_t) is symmetric with respect to the individuals, that is invariant by permutation of the individual indexes.

By a similar argument, the distribution of any exogenously given subset of m rating chains $(Y_{i_1,t},...,Y_{i_m,t})$ is symmetric with respect to the specific individuals introduced in the set. This property of symmetry is a condition of homogeneity of the population [see the discussion below of this homogeneity assumption].

In general the distribution of Y_{t+1} given the past ratings depends on the whole history $\underline{Y_t}$, and the joint process (Y_t) is not Markov[6]. The distribution of Y_{t+1} given $\underline{Y_t}$ can be summarized by a $K^n \times K^n$ transition matrix from Y_t to Y_{t+1}, whose elements depend on the whole past history $\underline{Y_{t-1}}$. After an appropriate reordering of the states $\{1,...,K\}^n$, this transition matrix is given by:

$$P_n = E\left[\overset{n}{\otimes}\Pi\left(Z_{t+1}\right) \mid \underline{Y_t}\right],$$

where $\overset{n}{\otimes}$ denotes n-fold Kronecker product[7].

iii) Finally both individual histories and factors can be observed up to time t. The available information set becomes $(\underline{Y_t},\underline{Z_t})$. Then the joint transition probabilities are derived by integrating out the future factor values only. The transition matrix is given by:

$$Q_n = E\left[\overset{n}{\otimes}\Pi\left(Z_{t+1}\right) \mid Z_t\right].$$

[6] However the process (Y_t) is Markov if the transition matrices (Π_t) are i.i.d., see Section 3.1 for a detailed discussion.

[7] If $A=[a_{ij}]$ and $B=[b_{ij}]$ denote $n\times m$ and $p\times q$ matrices, respectively, the Kronecker product $A\otimes B$ is a $np\times mq$ matrix with blocks of dimension $p\times q$ arranged on a $n\times m$ array, such that the block in position (i,j) is given by $[a_{ij}B]$.

The effect of this change in the information set will be discussed more carefully later on [see Section 2.3].

2.2 The homogeneity assumption

As noted above the property of symmetry is a condition of homogeneity of the population of individuals. It implies identical distributions for the individual histories, but also equidependence [see e.g. Frey, Mc Neil (2001), Gouriéroux, Monfort (2002)]. This equidependence property is a consequence of the common stochastic transition matrix for each date. The empirical relevance of the homogeneity assumption has to be discussed for the application to credit risk. For this purpose it is necessary to distinguish between retail credit and corporate bonds.

i) The retail credits include consumer credits, such as mortgages, classical consumption credits, revolving credits as well as over-the-counter credits to small and medium size firms. For such applications the number of borrowers is very large between 100000 and several millions. The practice for internal rating is to separate the population of borrowers into so-called "homogeneous" classes of risk where the individual risks can be considered as independent, identically distributed within the classes. For consumer credits the number of classes can be rather large (several hundreds), with classes including in general several thousands of individuals. The assumption of identical distributions and cross individual independence can be tested and these tests are the basis for determining the number of classes and their boundaries [see Gouriéroux, Jasiak (2003a) for a detailed presentation of the segmentation approach]. The homogeneity assumption considered in this paper extends the usual one in two respects. First it assumes identical dynamics for the individual ratings (not only identical marginal distributions for each Y_{it}). Second the condition of cross-independence is replaced by a condition of equidependence. From a practical point of view it corresponds to the approach proposed by Mc Kinsey for analyzing the risk of portfolios of retail credits[8].

ii) The situation is different when large corporates and the associated corporate bonds are considered. It is possible to classify these corporates

[8] without introducing necessarily observed macrovariables as underlying factors.

according to their rating, individual sector, ... and to expect a similar distribution of defaults in the medium run (1 year for S & P, 3 years for Banque de France). Indeed the ratings reported by the rating agencies such as Moody's, S & P, Fitch are derived according to this criterion. However it is not usual to check that the usual ratings can also be used for classifying defaults at other terms, and that they ensure equidependence. Typically, even if we could accept similar term structures of default (term structures of spreads, respectively) for IBM, General Motors, and Microsoft companies (resp. for the IBM, GM, Microsoft bonds) which have the same rating, the joint probability of IBM and Microsoft defaulting could be much higher. However it is important to note that both the professional and theoretical literatures often assume independence between defaults within a rating class for tractability reason, and in more recent literature "constant" default correlation [see e.g. Lucas (1995a), Duffie, Singleton (1999), Schonbucher (2000), Gordy, Heitfield (2002), de Servigny, Renault (2003)]. The assumption of equidependence is clearly a more flexible assumption than the two usual conditions, either independence, or constant default correlation, introduced in the literature.

2.3 Prediction and information

The transition matrices such as P_n, Q_n concern migration at horizon 1, that is short run migration. In practice the horizon of interest (for instance the investment or risk management horizon) can be different. In this section we discuss the term structure of migrations, which describes how the rating predictions depend on the horizon h. We distinguish the approach by prediction from the analysis by scenarios.

2.3.1 Prediction formulas

The dynamic MST specification can be used to analyze the rating predictions at different horizons. These predictions depend on the available information set, which can include either i) the lagged individual histories only, or ii) the lagged individual histories and the lagged factor values.

i) In the first case the predictive distribution of Y_{t+h} given $\underline{Y_t}$ is given by:

$$P_n^{(h)} = E\left[\overset{n}{\otimes}\Pi\left(Z_{t+1}\right)\overset{n}{\otimes}\Pi\left(Z_{t+2}\right)...\overset{n}{\otimes}\Pi\left(Z_{t+h}\right) \mid \underline{Y_t}\right]. \tag{1}$$

This matrix can be rewritten in terms of the transition matrix of the individual chains between t and $t+h$, which is denoted by $\Pi(t,t+h) = [\pi_{kl}(t,t+h)]$, and is defined by:

$$P\left[Y_{i,t+h}=l \mid Y_{i,t}=k,(Z_t)\right]=\pi_{kl}(t,t+h).$$

This transition matrix at horizon h is the product of h transition matrices at horizon 1:

$$\Pi(t,t+h)=\Pi_{t+1}\Pi_{t+2}...\Pi_{t+h}.$$

Then the predictive distribution of Y_{t+h} given $\underline{Y_t}$ becomes:

$$P_n^{(h)}=E\left[\overset{n}{\otimes}\Pi(t,t+h) \mid \underline{Y_t}\right].$$

ii) When the factors are observable up to time t, the predictive distribution of Y_{t+h} given $\underline{Y_t}, \underline{Z_t}$ becomes:

$$\begin{aligned} Q_n^{(h)} &= E\left[\overset{n}{\otimes}\Pi\left(Z_{t+1}\right)\overset{n}{\otimes}\Pi\left(Z_{t+2}\right)...\overset{n}{\otimes}\Pi\left(Z_{t+h}\right) \mid Z_t\right], \\ &= E\left[\overset{n}{\otimes}\Pi(t,t+h) \mid Z_t\right]. \end{aligned} \tag{2}$$

These transition matrices can be difficult to derive explicitly for two reasons. First the size of matrices $P_n^{(h)}$ and $Q_n^{(h)}$ can be very large; second the computation of the conditional expectations can be difficult. This explains the intensive use of Monte-Carlo technique to approximate the integrals (expectations). This explains also why we focus in practice on some summary statistics of the predictive distribution, such as the so-called migration correlations when the interest concerns a portfolio of firms (credits) [see Section 4].

2.3.2 Scenarios

The prediction formulas (1) and (2) have been derived for a given population and a future environment similar to the environment existing in the past. Both assumptions can be modified in the analysis by scenarios.

i) First we can fix the future path of the factor process. For instance let us assume that the future values $Z_{t+1}, Z_{t+2}, ...$ are known and fixed to the given identical level z. Then the transition at horizon h becomes: $P_n(z)^h$, with $P_n(z)=\overset{n}{\otimes}\Pi\left(z\right)$. In this approach the term structure of rating predictions will depend on the selected level z.

ii) It is also possible in the analysis by scenarios to up-date the population of interest. This approach is especially useful, when the state space includes an absorbing barrier, such as default in the rating analysis. Indeed, if the population of interest is kept fixed, all individuals will default asymptotically, and the population of firms (credits) which are still alive will systematically decrease[9]. This effect can be balanced in the scenario by assuming that at any date new firms are created to balance the failures [see Section 5.3 for a detailed discussion of such a rebalancing for statistical inference].

2.3.3 Factor observability for large population (large portfolios)

In some applications, such as the analysis of retail credit quality, the cross-sectional dimension n is often much larger than the time dimension T. In this situation it is interesting to study the limiting case $n \to \infty$.

Let us consider the historical data $(Y_{i,t})$, $i = 1, ..., n$. These data can be used to compute the migration counts $N_{kl,t}$, $k, l = 1, ..., K$, $t = 1, ..., T$, where $N_{kl,t}$ denotes the number of individuals migrating from k to l between $t-1$ and t, and the population structure per rating: $N_{k,t}$, $k = 1, ..., K$, $t = 1, ..., T$, where $N_{k,t}$ counts the number of individuals in state k at date t. It is easily checked[10] that the transition frequency:

$$\widehat{\pi}_{kl,t+1} = \frac{N_{kl,t+1}}{N_{k,t}},$$

tends to the theoretical transition probability at date $t+1$: $\pi_{kl,t+1}$, if n tends to infinity[11]. Thus, for a large population, the time dependent transition matrices Π_t, $t = 1, ...$ are known and also the factor Z_t, whenever the mapping: $Z \to \Pi(Z)$ is one-to-one. More explicitly let us denote by Π^- a (nonlinear) generalized inverse of function Π, the factor value is well approximated by:

$$\widehat{Z}_t = \Pi^-(\widehat{\Pi}_t). \tag{3}$$

[9] This assumption of fixed population of interest is imposed by the Basle Committee for the determination of the Value at Risk for a credit portfolio, that is of the capital required to hedge the risk included in the portfolio. This is the so-called "crystallisation" condition.

[10] by applying the Law of Large Numbers conditional on factor Z_{t+1}.

[11] This result is valid even if the state space includes an absorbing barrier. Indeed it is not necessary to balance "default" by new created firms, since the number of firms which are still alive is infinite at any date.

The formula above explains how to derive the factor Z_t from macrodata computed from individual histories. In particular, when n is large, it is no longer necessary to distinguish between the two information sets considered in Section 2.3.1.

3 Examples

This section describes different examples of MST models. We first consider the case of independent, identically distributed transition matrices. We describe some general properties of this specification and discuss the special case of transition matrices with beta distributions. The ordered polytomous model for transition matrices is reviewed in Section 3.2. This specification implies restrictions on the dynamics of transition matrices, which are pointed out. Finally, in Section 3.3 we consider reduced form modelling of the dynamics of the stochastic transition matrices as the Jacobi process.

3.1 Independent transition matrices

In this section we assume independent, identically distributed transition matrices.

Assumption A.1: The transition matrices (Π_t) [or factors (Z_t)] are independent, identically distributed (i.i.d.).

3.1.1 General properties

Under Assumption A.1, the joint transition of Y_{t+1} given $\underline{Y_t}$ is given by:

$$P\left(Y_{1,t+1} = y_{1,t+1}, ..., Y_{n,t+1} = y_{n,t+1} \mid \underline{Y_t}\right) = E\left(\prod_{k,l=1}^{K} \pi_{kl,t+1}^{n_{kl,t+1}}\right),$$

and depends on the individual histories by means of the state indicators for dates t and $t+1$ only[12]. We deduce the property below:

[12] In this formula the expectation is taken with respect to the stochastic transitions $(\pi_{kl,t+1})$, and the current and lagged state indicators are summarized in the observed counts $n_{kl,t+1}$.

Proposition 2: For a Markov stochastic transition model with i.i.d. factors (Z_t) [or (Π_t)], the joint process (Y_t) is an homogeneous Markov process.

The state space of this Markov process is $\{1, ..., K\}^n$ and, after appropriate ordering of the states, its transition matrix can be written as:

$$P_n = E\left[\overset{n}{\otimes}\Pi\left(Z_t\right)\right].$$

The homogeneity of the Markov process means that the migration intensity is independent of the term[13]. Thus Assumption A.1 implies a flat term structure of migration intensities.

By a similar argument, any given subset of m rating chains $(Y_{i_1,t}, ..., Y_{i_m,t})$ is Markov, with a transition matrix which depends on the size m, but not on the specific firms introduced in the set. For instance, for $m = 1$, the individual history (Y_{it}) of individual i is a Markov process with state space $\{1, ..., K\}$ and transition matrix $P_1 = E\left[\Pi(Z_t)\right]$, independent of the selected individual i. For $m = 2$ the bivariate individual histories $(Y_{i,t}, Y_{j,t})$ of the given pair of individuals (i, j) is a Markov process with state space $\{1, ..., K\}^2$ and transition matrix $P_2 = E\left[\Pi(Z_t) \otimes \Pi(Z_t)\right]$, independent of the selected pair of individuals (i, j).

Example 1: Let us consider the two state case $K = 2$, with state 2 as an absorbing barrier. The transition matrix can be written as:

$$\Pi_t = \begin{bmatrix} 1 - \pi_{12}(Z_t) & \pi_{12}(Z_t) \\ 0 & 1 \end{bmatrix}.$$

This specification is the basis for default analysis. The states are: state 1, if the credit (firm) is still alive, state 2, if it defaulted before time t. In this framework the non-vanishing joint transition probabilities take the forms $E\left[(1 - \pi_{12}(Z_t))^{n_{11,t+1}}\, \pi_{12}(Z_t)^{n_{12,t+1}}\right]$, where $n_{11,t+1}$ [resp. $n_{12,t+1}$] counts the credits (firms) staying alive between t and $t + 1$ [resp. defaulting between t and $t + 1$]. Up to a constant factor, this expression is the expectation of the elementary probability of a binomial distribution with stochastic parameter $\pi_{12}(Z_t)$, defining the default probability [see e.g. Schonbucher (2000), Gordy, Heitfield (2002) for similar expressions in the analysis of default correlation].

[13] Let $P_n^{(h)}$ denote the transition matrix of Markov process (Y_t) at horizon h. The migration intensity at horizon h is defined as the matrix $\Lambda_n^{(h)}$ such that $P_n^{(h)} = \exp\left(-h\Lambda_n^{(h)}\right)$. By the Markov property $P_n^{(h)} = (P_n)^h$, and thus the migration intensity $\Lambda_n^{(h)} = \Lambda_n$ is independent of the horizon h.

3.1.2 Transition matrices with beta distributions

Any row of the transition matrix $\pi_{[k],t} = (\pi_{kl,t},\ l = 1, ..., K)$ defines a stochastic discrete probability distribution. The beta model assumes independent transition matrices, with independent rows, following beta distributions[14].

Assumption A.2: Beta model

i) The rows $\pi_{[k],t},\ \ k = 1, ..., K,\ t = 1, ..., T$ are independent;

ii) $\pi_{[k],t}$ follows the beta distribution $Be\left(\alpha_k, \gamma_k\right)$, where $\alpha_k = (\alpha_{kl},\ l = 1, ..., K)$ is a deterministic discrete distribution: $\alpha_{kl} \geq 0$, $\forall k, l$, $\sum_{l=1}^{K} \alpha_{kl} = 1$, $\forall k$, and γ_k, $k = 1, ..., K$, are nonnegative parameters.

Let us recall that the beta distribution $Be\left(\alpha_k, \gamma_k\right)$ admits the density:

$$f_k(\pi_{[k]}) = \frac{\Gamma\left(\gamma_k\right)}{\prod_{l=1}^{K} \Gamma\left(\alpha_{kl}\gamma_k\right)} \prod_{l=1}^{K} \left(\pi_{kl}\right)^{\alpha_{kl}\gamma_k - 1}, \tag{4}$$

and a support corresponding to discrete probability distributions on the state space $\{1, ..., K\}$. The mean of the beta distribution is:

$$E\left(\pi_{[k],t}\right) = \alpha_k, \tag{5}$$

whereas its variance-covariance matrix is:

$$V\left(\pi_{[k],t}\right) = \frac{diag(\alpha_k) - \alpha_k \alpha_k^{'}}{\gamma_k + 1}. \tag{6}$$

In particular, migration probabilities to different rating classes always feature negative correlation, since $Cov(\pi_{kl,t}, \pi_{kl^*,t}) = -\alpha_{kl}\alpha_{kl^*}/(1 + \gamma_k) \leq 0$. Up to the scale factor $\left(\gamma_k + 1\right)^{-1}$, the variance-covariance matrix $V\left(\pi_{k,t}\right)$ coincides with the standard variance-covariance matrix of a multinomial distribution. The γ_k parameter can be seen as a state dependent volatility parameter: the smaller is parameter γ_k, the larger is the variance[15]. In the limiting case $\gamma_k = 0$, the beta distribution coincides with the multinomial distribution $\mathcal{M}\left(K, \alpha_{k1}, ..., \alpha_{kK}\right)$, whereas $\pi_{k,t}$ becomes deterministic when $\gamma_k = +\infty$.

[14] The beta distribution is the basic specification used in probability and statistics for the law of a stochastic discrete distribution.

[15] Another interpretation of parameter γ_k in terms of migration correlation is provided in Section 4.2.

The beta specification is easy to implement both for prediction and estimation purpose. For instance the joint transition probabilities are given by:

$$\begin{aligned}
& P\left(Y_{1,t+1}=y_{1,t+1},...,Y_{n,t+1}=y_{n,t+1} \mid Y_t\right) \\
= \; & E\left(\prod_{k,l=1}^{n} \pi_{kl,t+1}^{n_{kl,t+1}}\right) \\
= \; & \prod_{k=1}^{K} \frac{\Gamma\left(\gamma_k\right)}{\Gamma\left(\gamma_k+n_{k,t}\right)} \prod_{l=1}^{K} \frac{\Gamma\left(\alpha_{kl}\gamma_k+n_{kl,t+1}\right)}{\Gamma\left(\alpha_{kl}\gamma_k\right)}.
\end{aligned}$$

3.2 Ordered polytomous model

A specification which is often suggested by the agencies proposing measures of credit risk [see e.g. Gupton, Finger, Bhatia (1997), Crouhy, Galai, Mark (2000), Bangia et alii (2002), Albanese et alii (2003), Feng et alii (2003)] is the ordered polytomous model. The idea is to introduce an unobservable quantitative score from which the qualitative ratings are computed. Such a latent grade is sometimes computed regularly by the rating specialists, especially for internal ratings of retail credits [see Gouriéroux, Jasiak (2003)a]. Generally it is confidential and has to be considered as unobservable. In other approaches based on Merton's model [Merton (1974)], this latent grade is defined as the ratio of asset value and liabilities. It is also unobservable, whenever a detailed balance sheet of the firm is not available.

More specifically, let us denote by s_{it} the underlying quantitative score for corporate i at date t. Let us assume that the conditional distribution of s_{it} given the past depends on a factor Z_t (which can be multidimensional) and on the previous rating $Y_{i,t-1}$, and is such that:

$$s_{it} = \alpha_k + \beta_k' Z_t + \sigma_k \varepsilon_{it}, \tag{7}$$

if $Y_{i,t-1} = k$, where (ε_{it}) are iid variables with cdf G, and the common factor (Z_t) is independent of (ε_{it}). Thus three parameters are introduced for each initial rating class: α_k represents a level effect, the components of β_k define the sensitivities, whereas σ_k corresponds to a volatility effect.

The econometric model (7) includes several explanatory variables that are indicators of the lagged rating class and the cross effects between these indicators and the different factors. The specification focuses on the omitted time dependent factors. It implicitly assumes that the fixed individual effects have already been taken into account in the construction of the score.

Let us finally assume that the qualitative rating at date t is deduced by discretizing the underlying quantitative score:

$$Y_{i,t} = l, \text{ iff } \quad c_l \leq s_{it} < c_{l+1}, \tag{8}$$

where $c_1 = -\infty < c_2 < ... < c_K < c_{K+1} = +\infty$ are fixed (unknown) thresholds. This specification satisfies a MST model, with stochastic transition probabilities given by:

$$\begin{aligned}
\pi_{kl,t} &= P\left[Y_{it} = l \mid Y_{i,t-1} = k, Z_t\right] \\
&= P\left[c_l \leq \alpha_k + \beta_k^{'} Z_t + \sigma_k \varepsilon_{it} < c_{l+1} \mid Y_{i,t-1} = k, Z_t\right] \\
&= P\left[\frac{c_l - \alpha_k - \beta_k^{'} Z_t}{\sigma_k} < \varepsilon_{it} < \frac{c_{l+1} - \alpha_k - \beta_k^{'} Z_t}{\sigma_k} \mid Z_t\right] \\
&= G\left(\frac{c_{l+1} - \alpha_k - \beta_k^{'} Z_t}{\sigma_k}\right) - G\left(\frac{c_l - \alpha_k - \beta_k^{'} Z_t}{\sigma_k}\right), \quad k, l = 1, ..., K.
\end{aligned} \tag{9}$$

We get an ordered polytomous model for each row, with a common latent factor Z_t. When factor Z_t is unobservable the model has to be completed by specifying the factor dynamics.

The parameters of the ordered polytomous model (9) and of the factor dynamics are not identifiable, when the migration probabilities $\pi_{kl,t}$ are known. First the factor is defined up to an invertible affine transformation; thus we can always assume:

identifying constraints on the factor dynamics:
$E\left(Z_t\right) = 0, \quad V\left(Z_t\right) = Id.$

Second other identifiability problems are due to the partial observability of the quantitative score. The same migration probabilities can be obtained with an affine transformation of the quantitative score, which can depend on the initial rating and appropriate change of the thresholds. The identifiable functions of the parameters are:

$$\begin{aligned}
\left(c_l - \alpha_k\right)/\sigma_k, \ k &= 1, ..., K, \ l = 2, ..., K, \\
\beta_k/\sigma_k, \ k &= 1, ..., K,
\end{aligned}$$

or equivalently:

$$(c_2 - \alpha_k)/\sigma_k,\ \ (c_3 - c_2)/\sigma_k,\ ..., (c_K - c_2)/\sigma_k,\ \ \beta_k/\sigma_k,\ \ k = 1, ..., K.$$

A first identification restriction is $c_2 = 0$, which provides the new set of transformed parameters:

$$\alpha_k/\sigma_k,\ c_3/\sigma_k,\ ..., c_K/\sigma_k,\ \beta_k/\sigma_k,\ k = 1, ..., K.$$

It is easily checked that the remaining parameters are identified up to a scale factor. Thus an additional identifying restriction is: $\sigma_1^2 = 1$. Thus we assume:

identifying restrictions for partial observability:
$c_2 = 0, \quad \sigma_1^2 = 1.$

i) The model reduces to a probit model if the error terms (ε_{it}) follow a standard Gaussian distribution as proposed in CreditMetrics, for instance. The migration probabilities become:

$$\pi_{kl,t} = \Phi\left(\frac{c_{l+1} - \alpha_k - \beta_k' Z_t}{\sigma_k}\right) - \Phi\left(\frac{c_l - \alpha_k - \beta_k' Z_t}{\sigma_k}\right),$$

where Φ denotes the cdf of the standard normal.

ii) When the error terms are $\varepsilon_{it} = \log u_{it}$, with (u_{it}) following an exponential distribution, the cdf G corresponds to a Gompertz distribution: $G(x) = 1 - \exp(-e^x)$. The migration probabilities are given by:

$$\pi_{kl,t} = \exp\left[-\exp\left(\frac{c_l - \alpha_k - \beta_k' Z_t}{\sigma_k}\right)\right] - \exp\left[-\exp\left(\frac{c_{l+1} - \alpha_k - \beta_k' Z_t}{\sigma_k}\right)\right],$$

This model is a multistate extension of the two state Cox model with stochastic intensity, usually considered for corporate bond pricing [see e.g. Lando (1998), Gouriéroux, Monfort, Polimenis (2003)]. Indeed the model is equivalent to:

$$Y_{i,t} = l, \ \text{ if } \ d_{k,l,t} \leq u_{it} \leq d_{k,l+1,t},$$

where $d_{k,l,t} = \exp\left(\frac{c_l - \alpha_k - \beta_k' Z_t}{\sigma_k}\right)$, if $Y_{i,t-1} = k$. Thus the transitions are induced by an exponential variable crossing a grid of stochastic thresholds, where the grid depends on the starting grade.

3.3 Reduced form models

It is also possible to introduce directly a dynamics for the transition matrices without referring to any structural latent variables. The aim is to extend the basic i.i.d. beta model described in Section 3.1 by allowing for serial dependence of the successive transition matrices. Two extensions are presented below.

3.3.1 Jacobi specification

Any row of the (stochastic) transition matrix, such as the first one, defines a (stochastic) discrete probability distribution on the state space $\{1, ..., K\}$. Thus, when t varies, we get a stochastic process with values on the set of discrete distributions. The Jacobi process has been introduced to specify the dynamics of such a stochastic discrete probability distribution [see Gouriéroux, Jasiak (2003)b].

A Jacobi specification for the stochastic transitions assumes:

i) the rows $[\pi_{kl,t},\ l = 1, ..., K,\ t = 1, ..., T]$, $k = 1, ..., K$, are independent stochastic processes;

ii) any row $[\pi_{kl,t},\ l = 1, ..., K,\ t = 1, ..., T]$ corresponds to the discrete time observations of a continuous time multivariate Jacobi process, which satisfies the diffusion system:

$$d\pi_{kl,t} = b_k(\pi_{kl,t} - a_{kl})dt + \sqrt{g_k \pi_{kl,t}} dW_{kl,t} - \pi_{kl,t} \sum_{m=1}^{K} \sqrt{g_k \pi_{km,t}} dW_{km,t}, \quad l = 1, ..., K,$$

where $(W_{km,t})$, $k, m = 1, ..., K$, are independent Brownian motions, and the parameters satisfy the constraints $b_k < 0, g_k > 0$, $\sum_{l=1}^{K} a_{kl} = 1, \forall k$, $a_{kl} > 0, \forall k, l$.

The drifts of the diffusions suggest that the processes $[\pi_{kl,t},\ l = 1, ..., K,\ t = 1, ..., T]$ feature a mean-reverting dynamics, with equilibrium levels $a_{kl}, l = 1, ..., K$, and mean-reverting parameters b_k, $k = 1, ..., K$. The serial dependence of the stochastic transition matrices (Π_t) is controlled by the mean-reverting parameters b_k. The parameters g_k can be interpreted either as volatility parameters, or as smoothing parameters. In particular, if these parameters tend to infinity, the process $\pi_{kl,t}$ tends to a pure jump process. The restrictions on the parameters ensure that each row is a stationary process, with beta stationary distribution. For instance in the bivariate case the distribution is $Be(\alpha_k, \gamma_k)$, where $\alpha_k = (a_{k1}, a_{k2})$, $\gamma_k = -2b_k/g_k$.

3.3.2 Logistic autoregression

Serial dependence can also be directly introduced by considering Gaussian vector autoregressions applied to transformed transition probabilities. For instance let us consider the two state case $K = 2$. The transition matrix is characterized by the transition probabilities $\pi_{11,t}, \pi_{22,t}$. A logistic autoregression can be introduced for $\pi_{11,t}, \pi_{22,t}$. It is specified as:

$$\begin{aligned}
\log \frac{\pi_{11,t}}{1-\pi_{11,t}} &= c_1 + \varphi_{11} \log \frac{\pi_{11,t-1}}{1-\pi_{11,t-1}} + \varphi_{12} \log \frac{\pi_{22,t-1}}{1-\pi_{22,t-1}} + \varepsilon_{1t}, \\
\log \frac{\pi_{22,t}}{1-\pi_{22,t}} &= c_2 + \varphi_{21} \log \frac{\pi_{11,t-1}}{1-\pi_{11,t-1}} + \varphi_{22} \log \frac{\pi_{22,t-1}}{1-\pi_{22,t-1}} + \varepsilon_{2t},
\end{aligned}$$

where $(\varepsilon_{1t}, \varepsilon_{2t})$ is a Gaussian white noise with variance-covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$. Such a logistic transformation of transition probabilities before introducing the Gaussian autoregression is for instance suggested in the Mc Kinsey methodology[16].

However the approach by logistic autoregressions is difficult to extend to a larger number of states. Indeed there is no general agreement on a multivariate one-to-one transformation which associates with probabilities $(\pi_{1,t}, ..., \pi_{K-1,t})$, say, constrained by $0 \leq \pi_{1,t} \leq 1$, ..., $0 \leq \pi_{K-1,t} \leq 1$, $0 \leq 1 - \sum_{k=1}^{K-1} \pi_{k,t} \leq 1$, an unconstrained vector of $\mathbb{R}^{K-1}$.

4 Migration correlation

In order to analyse the default risk in a credit portfolio, it is important to take into account carefully the simultaneous rating migrations of different firms in the same direction, such as joint up- or downgrades. The tendency to a common rating migration is called migration correlation [see e.g. Lucas (1995a, b), Bahar, Nagpal (2001), de Servigny, Renault (2002)]. It extends

[16] But the logistic transformation for the lagged variables has been omitted in the Mc Kinsey methodology.

the concept of default correlation, corresponding to the two state case with default absorbing barrier.

It is important to realize that the standard specification with constant (deterministic) transition matrices does not feature migration correlation. In our framework, migration correlation between different rating classes is introduced by means of the stochastic transition matrix, which is a common (multivariate) factor across firms. In some sense it is not possible to distinguish between migration correlation and (stochastic) change in migration probabilities [see Lucas (1995a) p82]. More precisely the basic model (see Section 2) considers the conditional distribution of firm ratings given the sequence of transition matrices and assumes the conditional independence between firms. A dependence between rating dynamics is deduced when the stochastic transition matrices are integrated out, which creates migration dependence.

We first consider the case of i.i.d. transition matrices corresponding to a flat term structure of migration intensity. As seen in Section 3.1, the integration of the unobservable future values of the factor does not destroy the Markov property and the dynamics does not depend on the selected information set. In Section 4.1 we define precisely the notions of joint bivariate transition and of migration correlation, and explain how they can be displayed in well-chosen matrices. The definitions are illustrated in Section 4.2, where the migration correlations are computed for the beta specification, and in Section 4.3 for the ordered polytomous model. The definition of migration correlation has to be reconsidered when the transition matrices are serially dependent. This is done in Section 4.4, where the importance of the selected information set is emphasized.

4.1 Definition in the i.i.d. case

Let us consider two firms, whose rating histories are described by the chains $(Y_{i,t})$ and $(Y_{j,t})$, respectively, following a MST specification with i.i.d. transition matrices. In Section 3.1 it has been proved that the bivariate process $(Y_{i,t}, Y_{j,t})$ is still a Markov process. The bivariate joint transition is characterized by:

$$p_{kk^*,ll^*} = P\left[Y_{i,t+1} = k^*, Y_{j,t+1} = l^* \mid Y_{i,t} = k, Y_{j,t} = l\right] = E\left[\pi_{kk^*,t}\pi_{ll^*,t}\right]. \quad (10)$$

It defines a $K^2 \times K^2$ square matrix of joint transition probabilities. Similarly we get:

$$P\left[Y_{i,t+1} = k^* \mid Y_{i,t} = k, Y_{j,t} = l\right] = E\left[\pi_{kk^*,t}\right]. \qquad (11)$$

Migration correlation is defined in terms of the conditional correlation of the individual rating indicators[17]:

$$\rho_{kk^*,ll^*} = corr\left(\mathbb{I}_{Y_{i,t+1}=k^*}, \mathbb{I}_{Y_{j,t+1}=l^*} \mid Y_{i,t} = k, Y_{j,t} = l\right),$$

where $\mathbb{I}_{Y_{i,t+1}=k^*} = 1$, if $Y_{i,t+1} = k^*$, $= 0$, otherwise. The migration correlations can be written in terms of the underlying stochastic transition probabilities. From (10) and (11) we get:

$$\begin{aligned} \rho_{kk^*,ll^*} &= \frac{cov\left(\mathbb{I}_{Y_{i,t+1}=k^*}, \mathbb{I}_{Y_{j,t+1}=l^*} \mid Y_{i,t} = k, Y_{j,t} = l\right)}{V\left(\mathbb{I}_{Y_{i,t+1}=k^*} \mid Y_{i,t} = k, Y_{j,t} = l\right)^{1/2} V\left(\mathbb{I}_{Y_{j,t+1}=l^*} \mid Y_{i,t} = k, Y_{j,t} = l\right)^{1/2}} \\ &= \frac{cov\left(\pi_{kk^*,t}, \pi_{ll^*,t}\right)}{\left[E\pi_{kk^*,t}\left(1 - E\pi_{kk^*,t}\right)\right]^{1/2}\left[E\pi_{ll^*,t}\left(1 - E\pi_{ll^*,t}\right)\right]^{1/2}} \qquad (12) \\ &= corr\left(\pi_{kk^*,t}, \pi_{ll^*,t}\right)\left[\frac{V\left(\pi_{kk^*,t}\right)}{E\pi_{kk^*,t}\left(1 - E\pi_{kk^*,t}\right)}\right]^{1/2}\left[\frac{V\left(\pi_{ll^*,t}\right)}{E\pi_{ll^*,t}\left(1 - E\pi_{ll^*,t}\right)}\right]^{1/2}. \qquad (13) \end{aligned}$$

Migration correlation between firms i and j depends on their current and future ratings only, not on their names i and j. Any migration correlation involves three components: the first one $corr\left(\pi_{kk^*,t}, \pi_{ll^*,t}\right)$ measures the link between the underlying stochastic transition probabilities. The two other ones of the type $\left[V\left(\pi_{kk^*,t}\right)/E\pi_{kk^*,t}\left(1 - E\pi_{kk^*,t}\right)\right]^{1/2}$ are between 0 and 1. They take value 1 if the distribution of the transition probability weights 0 and 1 only, and take value 0 if the transition probability is constant[18].

There are as many different migration correlations as unordered pairs $(k, k^*), (l, l^*)$, that is $K^2\left(1 + K^2\right)/2$ different correlations. The migration correlations ρ_{kk^*,ll^*} actually involve cross-correlations between different firms i and j. In particular the diagonal terms $k = k^*, l = l^*$ are not necessarily equal to 1. The migration correlations are linearly dependent. They satisfy restrictions such as: $\sum_{k^*} \rho_{kk^*,ll^*}\left[E\pi_{kk^*,t}\left(1 - E\pi_{kk^*,t}\right)\right]^{1/2} = 0$, $\forall k, l, l^*$, due

[17] By a similar approach we can define migration correlations at any horizon h by considering the correlation between the rating indicators $\mathbb{I}_{Y_{i,t+h}=k^*}$ and $\mathbb{I}_{Y_{j,t+h}=l^*}$ conditional on $Y_{i,t} = k, Y_{j,t} = l$.

[18] See below for an interpretation of these terms.

to the unit mass restrictions on transition probabilities. In particular they cannot be of the same sign. Among these migration correlations some are more appealing for practitioners, especially those which involve migrations of the firms by one rating tick. For instance we can consider correlations between upgrades. If the initial state is (k, l) the correlation is given by:

$$corr\left(\mathbb{I}_{Y_{i,t+1}=k+1}, \mathbb{I}_{Y_{j,t+1}=l+1} \mid Y_{i,t} = k, Y_{j,t} = l\right), \qquad k, l = 1, ..., K-1.$$

We get a $(K-1) \times (K-1)$ symmetric matrix of up-up migration correlations indexed by the current ratings. Similarly we can define the down-down migration correlation matrix, which is a $(K-1) \times (K-1)$ symmetric matrix with elements:

$$corr\left(\mathbb{I}_{Y_{i,t+1}=k-1}, \mathbb{I}_{Y_{j,t+1}=l-1} \mid Y_{i,t} = k, Y_{j,t} = l\right), \quad k, l = 2, ..., K.$$

Finally we can consider rating migrations in opposite directions. The up-down migration correlation matrix has the elements:

$$corr\left(\mathbb{I}_{Y_{i,t+1}=k+1}, \mathbb{I}_{Y_{j,t+1}=l-1} \mid Y_{i,t} = k, Y_{j,t} = l\right).$$

This matrix is not symmetric in general. Other interesting migration correlation matrices are those involving identical starting and identical final ratings for the firms:

$$corr\left(\mathbb{I}_{Y_{i,t+1}=k^*}, \mathbb{I}_{Y_{j,t+1}=k^*} \mid Y_{i,t} = k, Y_{j,t} = k\right), \quad k, k^* = 1, ..., K.$$

We get a $K \times K$ matrix indexed by the starting and final ratings of the firms. From the migration correlation analysis equation (13), we deduce that these correlations are equal to $V\left(\pi_{kk^*,t}\right)/E\pi_{kk^*,t}\left(1 - E\pi_{kk^*,t}\right)$, providing an interpretation for the second and third terms in equation (13).

Example 1 (cont.): Let us consider the two state case $K = 2$ with default absorbing barrier $(k = 2)$. Then the transition matrix admits the form:

$$\Pi_t = \begin{pmatrix} 1 - \pi_{12,t} & \pi_{12,t} \\ 0 & 1 \end{pmatrix},$$

where $\pi_{12,t}$ is the default probability between $t-1$ and t.

The migration correlation matrices are scalar. The up-up and up-down correlations are equal to 0, since the default alternative is an absorbing barrier. The down-down correlation is given by:

$$\begin{aligned} & corr\left(\mathbb{I}_{Y_{i,t+1}=2}, \mathbb{I}_{Y_{j,t+1}=2} \mid Y_{i,t}=1, Y_{j,t}=1\right) \\ = & \frac{V\left(\pi_{12,t}\right)}{E\pi_{12,t}\left(1-E\pi_{12,t}\right)}, \end{aligned}$$

which corresponds to the usual definition of the instantaneous default correlation. The whole migration correlation matrix $R = \left(\rho_{kk^*,ll^*}\right)$ is:

$$R = \frac{V\left(\pi_{12,t}\right)}{E\pi_{12,t}\left(1-E\pi_{12,t}\right)} \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where the rating pairs are ranked as $(1,1), (1,2), (2,1), (2,2)$. Since default is an absorbing barrier, the only non vanishing migration correlations have state $(1,1)$ as initial state. These migration correlations are of the same size in absolute value. Migration correlations involving migrations to the same (opposite, respectively) rating classes are positive (negative, respectively).

Finally remark that the migration correlations are related to the joint migration probabilities by the formula:

$$\begin{aligned} & P\left[Y_{i,t+1}=k^*, Y_{j,t+1}=l^* \mid Y_{i,t}=k, Y_{j,t}=l\right] \\ = & P\left[Y_{i,t+1}=k^* \mid Y_{i,t}=k\right] \cdot P\left[Y_{j,t+1}=l^* \mid Y_{i,t}=k\right] \\ & +\rho_{kk^*,ll^*}\left[E\pi_{kk^*,t}\left(1-E\pi_{kk^*,t}\right)\right]^{1/2}\left[E\pi_{ll^*,t}\left(1-E\pi_{ll^*,t}\right)\right]^{1/2}. \end{aligned} \quad (14)$$

The second term in equation (14) describes the effect of migration correlation on the joint migration probability. The latter is equal to the product of the marginal migration probabilities of the firms if and only if the migration correlation is zero.

4.2 Migration correlation in the beta model

Let us now derive the migration correlations in the beta model introduced in Section 3.1.2. The up-up and down-down migration correlations at horizon

1 are identical, diagonal matrices [see Appendix 1], where:

$$\begin{aligned}
& corr\left(\mathbb{I}_{Y_{i,t+1}=k+1}, \mathbb{I}_{Y_{j,t+1}=l+1} \mid Y_{i,t}=k, Y_{j,t}=l\right) \\
= & \; corr\left(\mathbb{I}_{Y_{i,t+1}=k-1}, \mathbb{I}_{Y_{j,t+1}=l-1} \mid Y_{i,t}=k, Y_{j,t}=l\right) \\
= & \left\{\begin{array}{ll} \frac{1}{1+\gamma_k}, & \text{if } k=l, \\ 0, & \text{otherwise.} \end{array}\right.
\end{aligned}$$

The up-down migration correlation matrix is diagonal too with elements:

$$\begin{aligned}
& corr\left(\mathbb{I}_{Y_{i,t+1}=k+1}, \mathbb{I}_{Y_{j,t+1}=l-1} \mid Y_{i,t}=k, Y_{j,t}=l\right) \\
= & \left\{\begin{array}{ll} -\frac{1}{1+\gamma_k}\left[\frac{\alpha_{k,k+1}}{1-\alpha_{k,k+1}} \frac{\alpha_{k,k-1}}{1-\alpha_{k,k-1}}\right]^{1/2}, & \text{if } k=l, \\ 0, & \text{otherwise.} \end{array}\right.
\end{aligned}$$

Finally, the migration correlations corresponding to identical initial and final grades are given by:

$$corr\left(\mathbb{I}_{Y_{i,t+1}=k^*}, \mathbb{I}_{Y_{j,t+1}=k^*} \mid Y_{i,t}=k, Y_{j,t}=k\right) = \frac{1}{1+\gamma_k}.$$

The assumptions of the basic beta migration model[19] imply restrictions on migration correlations. For instance, the migration correlations at horizon 1 are different from zero only when the two firms are currently in the same rating class. This is a consequence of the assumption of independent rows for the transition matrix. Moreover, when the firms are in the same rating class, the migration correlations to identical final ratings (for instance up-up or down-down migrations) are all equal, whatever the final rating. This is due to the special form of the variance-covariance matrix of the beta distribution. Note however that the model is sufficiently flexible to allow for different correlations according to the current state. More specifically the migration correlation in rating class k for migrations in similar directions is positive, and involves the parameter γ_k. The smaller is γ_k, the more important is the common factor, and the larger is the migration correlation. In particular, migration correlation achieves its maximal value 1, when $\gamma_k = 0$ (respectively, its minimal value 0, when $\gamma_k = \infty$). Therefore parameters γ_k admit a natural interpretation in terms of migration correlations of the different rating classes. Migration correlations in opposite directions are negative. Their absolute

[19] and more generally of any selected model.

value involves two terms, whose values are both between 0 and 1. The first term $1/(1+\gamma_k)$ is equal to the migration correlation in similar directions. The second involves odd-ratios of the expected transition probabilities $\alpha_{k,k+1}$ and $\alpha_{k,k-1}$, and achieves its maximal value 1 when $\alpha_{k,k+1}+\alpha_{k,k-1}=1$, that is when firms in rating class k may feature only an upgrade or a downgrade by one tick. In particular, the absolute value of the migration correlation in opposite directions is not larger than the migration correlation in similar directions.

Migration correlations at horizons larger than 1 are in general not equal to zero, even when the current ratings are different. Indeed the rows of transition matrices at larger horizons are not independent, whereas they are independent at short term horizon.

4.3 Migration correlation in a factor ordered qualitative model

As an illustration, let us consider the multistate extension of the Cox intensity model introduced in Section 3.2 with a single i.i.d. factor, and the down-grade probability for a firm rated k:

$$\widetilde{\pi}_{k,t} = \sum_{k^* \geq k} \pi_{k,k^*,t} = \exp\left[-\exp\left(\frac{c_k - \alpha_k - \beta_k Z_t}{\sigma_k}\right)\right].$$

An analytic expression for the down-grade migration correlation for two firms rated k and l :

$$\widetilde{\rho}_{k,l} = \frac{cov\left(\widetilde{\pi}_{k,t}, \widetilde{\pi}_{l,t}\right)}{\left[E\widetilde{\pi}_{k,t}\left(1-E\widetilde{\pi}_{k,t}\right)\right]^{1/2}\left[E\widetilde{\pi}_{l,t}\left(1-E\widetilde{\pi}_{l,t}\right)\right]^{1/2}},$$

can be obtained in the special case $\beta_k/\sigma_k = 1$, for any k. Indeed we get:

$$\widetilde{\pi}_{k,t} = \exp\left[-\lambda_k \exp\left(-Z_t\right)\right],$$

where $\lambda_k = \exp\left(\frac{c_k - \alpha_k}{\sigma_k}\right)$. We deduce:

$$\widetilde{\rho}_{k,l} = \frac{\Psi\left(\lambda_k + \lambda_l\right) - \Psi\left(\lambda_k\right)\Psi\left(\lambda_l\right)}{\sqrt{\Psi\left(\lambda_k\right)\left[1-\Psi\left(\lambda_k\right)\right]}\sqrt{\Psi\left(\lambda_l\right)\left[1-\Psi\left(\lambda_l\right)\right]}},$$

where Ψ denotes the Laplace transform of $\exp(-Z_t)$: $\Psi(u) = E[\exp(-u\exp(-Z_t))]$. Equivalently we have:

$$\widetilde{\rho}_{k,l} = \frac{\Psi\left[\Psi^{-1}(\pi_k) + \Psi^{-1}(\pi_l)\right] - \pi_k\pi_l}{\sqrt{\pi_k(1-\pi_k)}\sqrt{\pi_l(1-\pi_l)}},$$

where π_k denotes the marginal probability of down-grade in the class of risk k. This value depends on the marginal migration rates and on the factor distribution (by means of the Laplace transform Ψ). Since function $(x,y) \to \Psi\left[\Psi^{-1}(x) + \Psi^{-1}(y)\right]$ corresponds to an Archimedean copula [see e.g. Joe (1997)], a more heterogeneous factor will imply a larger migration correlation.

It has been usual among practitioners to compare the value of the migration correlations to the level of a latent correlation corresponding to the model of the underlying score[20]. In this example we have:

$$s_{i,t} = \alpha_k + \beta_k Z_t + \sigma_k \varepsilon_{i,t},$$

and thus a possible measure of the latent correlation is:

$$corr_{kl}(s_{i,t}, s_{j,t}) = \frac{\beta_k\beta_l V(Z_t)}{\sqrt{\beta_k^2 V(Z_t) + \sigma_k^2 V(\varepsilon_{i,t})}\sqrt{\beta_l^2 V(Z_t) + \sigma_l^2 V(\varepsilon_{i,t})}}.$$

However it is important to note that a quantitative score is defined up to an increasing transformation. In the extended Cox model it was more natural to consider the transformed score:

$$s_{i,t}^* = \exp(\alpha_k + \beta_k Z_t)\, u_{i,t}^{\sigma_k},$$

in order to get the crossing of a grid of thresholds by an exponential variate. Therefore another latent correlation can be defined:

$$corr_{kl}(s_{i,t}^*, s_{j,t}^*) = \frac{cov\left(e^{\beta_k Z_t}, e^{\beta_l Z_t}\right)}{\sqrt{V\left(e^{\beta_k Z_t}\right) + \frac{V\left(u_{i,t}^{\sigma_k}\right)}{E\left(u_{i,t}^{\sigma_k}\right)^2} E\left(e^{2\beta_k Z_t}\right)}\sqrt{V\left(e^{\beta_l Z_t}\right) + \frac{V\left(u_{i,t}^{\sigma_l}\right)}{E\left(u_{i,t}^{\sigma_l}\right)^2} E\left(e^{2\beta_l Z_t}\right)}}.$$

This value can be very different from the value computed directly from the score s, which moreover is not directly observable, and is also very different

[20] This is due to a suggestion of the Basle Committee to apply an ordered probit model with a latent correlation equal to 0.25, and to the difficulty in determining the corresponding level of the observable default correlation.

from the value of the down-grade migration correlations $\widetilde{\rho}_{k,l}$ [see the discussion in Section 6.2]. To summarize, the latent quantitative score is defined up to an increasing transformation and there exist as many latent correlations as admissible choices of the transformation. Thus the notion of latent correlation has to be used with care.

4.4 Migration correlations for serially dependent transition matrices

Joint bivariate transition probabilities can also be derived for serially dependent transition matrices [see Sections 3.2, 3.3]. However their expressions depend on the selected information set. They correspond to the matrix $P_2^{(1)}$, if the individual histories only are known up to time t, to the matrix $Q_2^{(1)}$, if both the individual histories and the factors are known up to time t [see equations (1) and (2)]. These two matrices coincide for large portfolios ($n = \infty$) [see Section 2.3.3].

Moreover for serially dependent transition matrices the joint bivariate process $(Y_{i,t}, Y_{j,t})$, where (i, j) is a given pair of individuals, is no longer Markov. Thus the joint bivariate transitions will not depend on the past through Y_{t-1} only, and in practice will vary with the date t. Similar remarks apply to the associated migration correlation matrices, which also depend on the selected information set and vary in time.

It is interesting to replace the definitions of migration correlation given above in a static or a dynamic framework with respect to the existing literature. Indeed estimated migration correlations have been displayed in the professional and academic literature [see e.g. Lucas (1995a), Bahar, Nagpal (2001), de Servigny, Renault (2002)], but "without relying on a specific model driving transitions" [de Servigny, Renault (2002)]. Since they are considered constant over the whole period of estimation, they correspond intuitively to a flat term structure of migration intensity, that is implicitly i.i.d. stochastic transition matrices have been assumed.

5 Statistical inference

As noted before the stochastic migration model is a special case of multifactor model for panel data. The likelihood function or the observable conditional moments can involve multidimensional integrals of rather large dimension, which depend generally on the number of observation dates. In such a framework the standard maximum likelihood or GMM approaches are numerically intractable and can be replaced by simulation approaches as simulated maximum likelihood, simulated method of moments or indirect inference [see e.g. the surveys by Gouriéroux, Monfort (1995), Gouriéroux, Jasiak (2001)]. However the stochastic migration model presents some specificities, and special aspects of statistical inference have to be discussed. For ease of exposition we first consider the estimation problem for a stochastic migration model with i.i.d. transition matrices and explain how to estimate an ordered qualitative model with dynamic factor in the last subsection. In Section 5.1 we discuss the consistency properties of the maximum likelihood estimator according to the dimension n or T, which tends to infinity. It is explained why a large cross-sectional dimension is not sufficient to get consistency. Section 5.2 considers the estimation problem when only the aggregate structures per rating, not the individual migration transitions, are available. In Section 5.3 we explain how the asymptotic theory has to be modified, when the state space includes an absorbing barrier (as default in the example of credit risk), and discusses the limiting case of large homogeneous populations (large portfolios). Finally Section 5.4 discusses statistical inference for the ordered qualitative model with dynamic factor when the cross-sectional dimension is large.

5.1 Consistency of the ML estimator

As noted earlier in Section 3.1 the multivariate rating process $Y_t = (Y_{1,t}, ..., Y_{n,t})'$ is a Markov process, when the transition matrices (Π_t) are independent identically distributed, with a $K^n \times K^n$ transition matrix P_n. Therefore its distribution is defined by the transition:

$$p(y_{t+1} \mid y_t; \theta) = \underset{\theta}{E}\left[\prod_{k=1}^{K}\prod_{l=1}^{K} \pi_{kl,t+1}^{N_{kl,t+1}}\right], \tag{15}$$

where θ denotes the parameter characterizing the distribution of Π_t, for instance $\theta = (\alpha_{kl}, \gamma_k,\ k, l = 1, ..., K)$ for the beta model. The ML estimator is a solution of the optimization problem:

$$\widehat{\theta} = \arg\max_{\theta} \sum_{t=2}^{T} \log p\left(y_t \mid y_{t-1}; \theta\right). \tag{16}$$

The likelihood function depends on the observations by means of the aggregate migration counts $N_{kl,t}$, $k, l = 1, ..., K$, $t = 1, ..., T$, which constitutes a sufficient statistics for θ. The likelihood involves multidimensional integrals with dimension less or equal to $K(K-1)$; this dimension does not depend on the number of observations nT [21].

If n is fixed and T tends to infinity the general asymptotic theory for Markov chains can be applied [Anderson, Goodman (1957)]. In particular the ML estimator is consistent under regularity conditions including the assumption that the chain is recurrent, that is passes an infinite number of times by any admissible state (when T tends to infinity).

At the opposite, when n tends to infinity and T is fixed, the ML estimator of θ is not consistent[22]. This feature is easily understood if we consider the case $T = 1$. The sufficient statistics $N_{kl,1}$ can be used to compute the historical transition frequencies at date 1, that is $N_{kl,1}/N_{k,0}$, which tend to $\pi_{kl,1}$, when n tends to infinity (by the Law of Large Numbers applied conditional on the transition matrix Π_1). Therefore the transition matrix Π_1 is perfectly known. However the knowledge of Π_1, that is a single observation of the sequence of transition matrices, is not sufficient to identify the dynamics of Π_t, that is the parameter θ.

In summary the ML estimator is consistent for T tending to infinity, but not for n tending to infinity. The cross-sectional inconsistency of the estimator results from the cross-sectional equidependence, which does not allow the standard mixing conditions for the Law of Large Numbers to be satisfied. The remark on the non-consistency of the cross-sectional ML estimator of parameter θ is also valid when we consider other parameters of interest such as migration correlations[23] [see Section 4], or other estimation methods.

[21] This remark is no longer valid when the stochastic transition matrices feature serial dependence.

[22] See Gagliardini, Gouriéroux (2003) for a more detailed discussion.

[23] The ML estimator of the migration correlation is $\underset{\widehat{\theta}}{corr}\left(\pi_{kk^*,t}, \pi_{ll^*,t}\right) \cdot \left[\frac{\underset{\widehat{\theta}}{V}\left(\pi_{kk^*,t}\right)}{\underset{\widehat{\theta}}{E}\pi_{kk^*,t}\left(1-\underset{\widehat{\theta}}{E}\pi_{kk^*,t}\right)}\right]^{1/2} \left[\frac{\underset{\widehat{\theta}}{V}\left(\pi_{ll^*,t}\right)}{\underset{\widehat{\theta}}{E}\pi_{ll^*,t}\left(1-\underset{\widehat{\theta}}{E}\pi_{ll^*,t}\right)}\right]^{1/2}$, where θ is replaced by its ML estimator $\widehat{\theta}$.

5.2 Estimation from aggregated data on marginal structure per rating

In practice the data on transition counts $N_{kl,t+1}$ are not necessarily available, but the structure per rating $N_{k,t}$ are often provided by the rating agencies as Moody's, Standard & Poor's, ... [see e.g. Brady, Bos (2002)]. The parameter θ is easily estimated along the lines of McRae (1977). Indeed under the i.i.d. assumption on transition matrices, the counting process $(N_{k,t}, k = 1, ..., K)$ is also a Markov process [see Appendix 2]. In particular we get the moment restrictions:

$$\underset{t}{E}\left(N_{l,t+1}\right) = \sum_{k=1}^{K} P_{1,kl}(\theta) N_{k,t}, \tag{17}$$

where $P_{1,kl}(\theta)$ is the (k, l) element of the matrix $P_1(\theta) = \underset{\theta}{E}\left(\Pi_t\right)$. The parameter θ can be consistently estimated by nonlinear least squares when T tends to infinity (but not when n tends to infinity). The estimator is derived in two steps to account for the conditional heteroscedasticity of the counts (see Appendix 2). A first step estimator is derived by ordinary nonlinear least squares, that is by minimizing:

$$\widetilde{\theta} = \arg\min_{\theta} \sum_{t=2}^{T} \sum_{l=1}^{K} \left(N_{l,t+1} - \sum_{k=1}^{K} P_{1,kl}(\theta) N_{k,t} \right)^2 . \tag{18}$$

The OLS residuals are computed:

$$\widehat{u}_{l,t+1} = N_{l,t+1} - \sum_{k=1}^{K} P_{1,kl}(\widetilde{\theta}) N_{k,t}.$$

These residuals are used to apply a quasi-generalized nonlinear least squares approach. The second step estimator is:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{t=2}^{T} \sum_{l=1}^{K} \left(N_{t+1} - \sum_{k=1}^{K} P_{1,k}(\theta) N_{k,t} \right)' \Omega^{-1} \left(N_{t+1} - \sum_{k=1}^{K} P_{1,k}(\theta) N_{k,t} \right),$$

where $N_{t+1} = \left(N_{1,t+1}, ..., N_{K,t+1}\right)'$ is the vector of counts, $P_{1,k}(\theta)$ denotes the k-th column of matrix $P_1(\theta)$, and Ω is the matrix with (k, l) element $\sum_{t=2}^{T} \widehat{u}_{l,t} \widehat{u}_{k,t} / (T - 1)$.

5.3 The problem of absorbing barrier

The consistency of the ML estimator when T tends to infinity is satisfied under the recurrence condition. However in the credit risk applications, there exists an absorbing barrier (default state). Asymptotically all individuals are in default state and the migration parameters associated with a transitory phenomenon cannot be identified. For instance let us consider the regression (17). The observed structure $(N_{1,t}, ..., N_{K,t})$ tends to $(0, ..., 0, n)$, the explanatory variables are asymptotically colinear, and parameter θ is not identifiable.

A solution to recover the consistency of an estimator of θ for large T is to increase the size of the population in time in order to compensate the defaulted firms[24]. In a simple framework new individuals can be regularly introduced to get a fixed dimension of the population of alive individuals. When an individual i defaults at date t, it is replaced by a new one assigned randomly to a state $k = 1, ..., K-1$ according to a distribution $\mu_t = (\mu_{1,t}, ..., \mu_{K-1,t})$, say. The states occupied by this sequence of individuals define a process $\widetilde{Y}_t$, say, with state space $\{1, ..., K-1\}$. Conditionally to μ_t, Π_t, the process $\widetilde{Y}_t$ is a Markov process with a $(K-1) \times (K-1)$ transition matrix $\widetilde{\Pi}_t$. The elements of this matrix are:

$$\widetilde{\pi}_{kl,t} = \pi_{kl,t} + \pi_{kK,t}\mu_{l,t}, \qquad k, l = 1, ..., K-1.$$

The theory presented in Section 5.1, 5.2 can be applied to this transformed process $\left(\widetilde{Y}_t\right)$, whenever the matrices (Π_t, μ_t) are assumed i.i.d., with a specified parametric distribution. In this extension introduced to satisfy the recurrence condition and to get the consistency of the ML estimator, it is not possible to estimate the migration parameters without estimating jointly the parameters characterizing the process of population renewal[25].

It is interesting to particularize the limiting case of a very large population of individuals $n = \infty$. Despite default, the number of individuals alive at any date t is always infinite and some migrations between states can be observed even when t is very large. More precisely, by applying the cross-sectional argument, the transition matrices Π_t, $t = 1, ..., T$ are exactly known

[24] Such a regularly updated population is called "static pool" by Standard & Poor's [Brady, Bos (2002)].

[25] The procedure becomes simpler when the design of the population renewal is known by the econometrician.

(since they are consistently approximated by their empirical counterparts). Therefore the ML method can be applied to the observed factor Π_t, $t = 1, ..., T$. The ML estimators will be consistent when $n = \infty$ and T tends to infinity, even if there exist an absorbing barrier. In other words, in the case of an absorbing barrier the asymptotic bias of the ML estimator computed on a fixed population can be diminished by increasing the cross-sectional dimension.

5.4 Estimation of the ordered qualitative model with dynamic factor

Let us now explain how to estimate in a simple way the ordered qualitative model introduced in Section 3.2, when the cross-sectional dimension n is large and for instance the factors satisfy a Gaussian VAR process[26]:

$$Z_t = AZ_{t-1} + u_t,$$

where (u_t) is multivariate standard normal.

From (9) we deduce that:

$$\begin{aligned} \pi^*_{kl,t} &= \sum_{h<l} \pi_{kh,t} = P\left[Y_{i,t} < l \mid Y_{i,t-1} = k, Z_t\right] \\ &= G\left(\frac{c_l - \alpha_k - \beta_k^{'} Z_t}{\sigma_k}\right), \end{aligned}$$

or equivalently:

$$G^{-1}(\pi^*_{kl,t}) = \frac{c_l - \alpha_k}{\sigma_k} - \frac{1}{\sigma_k}\beta_k^{'} Z_t.$$

For a large cross-sectional dimension n ($n \to \infty$), the probability $\pi^*_{kl,t}$ is well-approximated by its cross-sectional sample counterpart $\widehat{\pi}^*_{kl,t}$, say; moreover we have $\sqrt{n}\left(\widehat{\pi}^*_{kl,t} - \pi^*_{kl,t}\right) \xrightarrow{d} N(0, \Omega_{kl,t})$, $t = 1, ..., T$, and the estimators corresponding to different dates are independent. By applying the δ-method along the lines initially proposed by Berkson (1944) for logit models with repeated observations, we get [see also Amemiya (1976)]:

[26] For expository purpose and the link with the state space representation we impose $Vu_t = Id$ instead of $VZ_t = Id$ as identifying constraint.

$$\left\{\begin{array}{l} G^{-1}(\widehat{\pi}^*_{kl,t}) \simeq \frac{c_l-\alpha_k}{\sigma_k} - \frac{1}{\sigma_k}\beta_k' Z_t + \widetilde{\Omega}^{1/2}_{kl,t} v_{kl,t}, \quad \forall k,l,t, \\ \text{where } Z_t = AZ_{t-1} + u_t, \end{array}\right.$$

where v_t and u_t are independent multidimensional Gaussian error terms. Thus, for large n, we get an approximated linear state space model, in which the macro-component corresponds to the transition equation and the micro-component to the measurement equation. This approximated model can be estimated by a standard Kalman filter, under the identification restrictions. This approach provides approximations of the microparameters α, c, β, σ, of the macro-parameters A and of the factor values. From general results on statistical inference for panel models with unobservable dynamic factors [Gourieroux, Monfort (2004)] it follows that:

i) the approximations of the factor values are $\sqrt{n}$-consistent;

ii) the estimators of the micro-parameters are $\sqrt{nT}$-consistent and asymptotically efficient;

iii) the estimator of the macro-parameters are $\sqrt{T}$-consistent.

To summarize, the large cross-sectional dimension is useful to introduce simple estimation approaches, based on the possibility to approximate the true transition matrices by their sample counterparts.

6 Application to migration data

6.1 The data set

Migration data are regularly reported by the rating agencies as Moody's and Standard and Poor's, or by some central banks as the Banque de France [see Foulcher, Gourieroux, Tiomo (2003) for a comparison of the main rating systems]. The data sets of the agencies concern rather large firms at the international level. The number of rated firms is about 10000 and reliable data are available since 1985. The rating is generally fixed by experts on the basis of information obtained at the moment of bond issuing, for instance. In these basis the proportion of missing data is rather large, which can induce selectivity bias in the empirical analysis.

In contrast the Banque de France collects yearly the balance sheets of all French firms. The balance sheets are used to construct a quantitative score explaining how default probability at 3 years depend on a set of financial ratios and individual characteristics. Then this score is discretized into rating classes. This data set has several advantages compared to the set of the agencies. First, it concerns about 180000 French firms, which allows to perform some analysis by size or industrial sectors without a too small number of observations. Second the formula for the econometric score can be followed, as well as the limiting thresholds, which define the rating classes [see Bardos et alii (2004)]. The rating procedure has been stable during the period of observation 1992-2003, which is not necessarily the case of the rating by expertise performed by the agencies, due to the change of experts or to their possible different rating behaviours during the phases of the business cycle (recessions and expansions). This explains why the application of this section will be performed on this more complete and reliable data set. We will focus on two industrial sectors corresponding to wholesale and retail trade, respectively. They concern about 30000 firms for each industrial sector [see Bardos et al (2004)], which are in general of small or medium size.

The Banque de France rating contains 8 classes of risk, denoted 0, 1, 2, ... 7 which represent the distance to default. For instance alternative "0" is for default, whereas alternative "7" represents the lowest risk, that is the usual AAA or Aaa of the rating agencies. The individual rating histories are aggregated to produce the transition matrices between rating classes for different years, categories and time horizons. For instance, such a 1-year transition matrix is given in Table 1, for year 2001 and the wholesale industrial sector.

[Table 1: Transition matrix in 2001 for the wholesale sector]

It is immediately seen that this matrix contains an additional column corresponding to the firms, which are not rated at the end of 2001. The firms are not rated due to missing data, which concern either the total balance sheet, or simply some financial ratios or characteristics introduced as explanatory variables in the underlying quantitative score. These missing data are mainly due to a lack of cooperation, which can be voluntary or not. The rate of missing data is rather large (between 10 and 30%) and larger than the rate generally observed for large firms (between 5 and 15%), which are obliged to report regularly some information concerning their balance sheet.

As the other rating agencies, the Banque de France does not report the row providing the transition from the not rated (NR) class to the other classes of risk. Therefore this type of matrix has to be transformed into a square transition matrix by assigning the non-rated companies among the other classes. They are usually assigned proportionally, which implicitly assumes the absence of selectivity bias [see all recent applied studies in the list of references for a similar approach]. It is seen on Table 1 that the NR rate has a tendency to increase when the quality of risk diminishes. This fact could be considered as an additional signal of bad risk, which would create a selectivity bias, or simply it can be due to the fact that providing information to the Banque de France is not a priority when the situation of the firm deteriorates[27]. An idea about the missing row can be obtained from Foulcher et al. (2003) where such a row looks like:

	7	6	5	4	3	2	1	0	NR
NR	2.5	2.5	2.0	2.0	1.4	1.1	0.2	0.5	87.8

This shows that the NR alternative is not an indicator of imminent default, which is in favour of a proportional assignment. This approach is followed in the rest of the section. The adjusted transition matrix corresponding to the matrix in Table 1 is provided in Table 2.

[Table 2: Adjusted transition matrix in 2001 for the wholesale sector]

As usual these transition matrices contains a lot of very small transition probabilities. The significant elements are essentially around the main principal diagonals, showing that the up- or down-grades are at most of one or two buckets during the year, for firms in a "standard" situation. These matrices are rather different from the transition matrices existing for large firms, for which the ratings are more stable. Typically in the S&P or Moody's data the three main diagonals only have significant elements. Finally note that the transition matrices are generally given without the last row corresponding to default as an initial state. Indeed default is (generally) an absorbing state and this row has zero elements except the last one which is equal to 1.

[27] More information on the type of missing data is given in Bardos et alii (2004).

6.2 I.I.D. transition matrices

Let us first consider a model with i.i.d. transition matrices. As mentioned in Section 3.1, the process of joint individual histories is an homogeneous Markov process. In this framework it is natural to compute the matrix of individual migration $P_1 = E\Pi_t$, and the matrix of joint migration for a pair of firms $P_2 = E(\Pi_t \otimes \Pi_t)$. The theoretical matrices are estimated by their sample counterparts, obtained by averaging on time the associated observed transition frequencies. The estimated matrices $\widehat{P}_1$ are given in Tables 3 and 4 for the wholesale sector and the retail trade sector, respectively:

[Table 3: Individual migration probabilities for the wholesale sector]

[Table 4: Individual migration probabilities for the retail trade sector]

Similarly we compute the estimated P_2 matrix and report some of its elements. In Table 5, we provide the joint up-grade probabilities in the wholesale sector:

[Table 5: Joint up-grade migration probabilities]

From these matrices we deduce the up-grade migration correlation in the wholesale sector for all pairs of initial states:

[Table 6: Up-grade migration correlations]

Similarly, we compute joint down-grade probabilities and down-grade correlations:

[Table 7: Down-grade migration probabilities]

[Table 8: Down-grade migration correlations]

Common features of the up- and down-grade correlation matrices can be observed.

i) The migration correlations are rather small, and typically much smaller than migration correlations reported by De Servigny, Renault (2002) from S&P data with two rating classes only, corresponding to investment and speculative grades. However these results are difficult to compare since they do not correspond to the same number of classes of risk, and it can be expected that the migration correlations will systematically diminish, when the partition becomes thinner[28].

[28]Indeed the correlations are conditional on the available information, that is the chosen segmentation and generally diminish when the information increases.

From a statistical point of view, it would be natural to test for the significance of such correlations. From a financial point of view such a test has to be avoided, since, even a small correlation can generate a risk, which can be artificially increased by a leverage effect created by the allocations in the credit portfolio.

ii) These correlations are generally non negative.

iii) They become smaller, when the two firms are more different, that is belong to distant classes of risk.

Finally, we can compute default correlations. They are reported in Tables 9 and 10 for the wholesale sector and the retail trade sector, respectively.

[Table 9: Default correlations in the wholesale sector]

[Table 10: Default correlations in the retail trade sector]

As for migration correlation, we observe rather small values, clearly much smaller than the values suggested by the Basle Committee (see Section 4.3 for the link between latent correlation and default correlation). As above this can be due to the partition into rating classes, which is neglected in the basic methodology suggested by the Basle Committee. The disaggregation of default correlation by rating classes can also have some other consequences. For instance it is known that in a large class of risk the default correlation is necessarily nonnegative [see e.g. Gourieroux, Monfort (2002) and Frey, Mc Neil (2003)]. By disaggregation, we can get subpopulation of smaller size and observe negative default correlations.

Anyway it is important to compare the default correlations across the rating classes with the estimated default correlation proposed by the Basle Committee. As mentioned in Section 4.3, the regulator suggests a probit model, with a latent correlation which is a function of the default probability π_k of the class to which the two firms belong [see the Basle Committee on Banking Supervision (2002)]. The relationship is:

$$\rho_k = 0.24 - 0.12 \frac{1 - \exp(-50\pi_k)}{1 - \exp(-50)}.$$

Since for a probit model the default correlation is given by:

$$\widetilde{\rho}_{k,k} = \frac{\int_{-\infty}^{\Phi^{-1}(\pi_k)} \int_{-\infty}^{\Phi^{-1}(\pi_k)} \frac{1}{2\pi\sqrt{1-\rho_k^2}} \exp\left(-\frac{1}{2\left(1-\rho_k^2\right)} \left(x^2 - 2\rho_k xy + y^2\right) \right) dxdy - \pi_k^2}{\pi_k \left(1 - \pi_k\right)},$$

we deduce the relationship between the default correlation and the marginal default probability proposed by the Basle Committee. This relationship is displayed in Figure 1 with a zoom on the range of observed default probabilities:

[Insert Figure 1: Default correlation as a function of default probability by Basle Committee]

It can be directly compared with the relationships estimated on the Banque de France data for the wholesale and retail trade sectors:

[Insert Figure 2: Default correlation as a function of default probability for two sectors]

It is seen that the estimated default correlations are systematically smaller than the values suggested by the regulator. Nevertheless they feature the same type of monotonic dependence with respect to the marginal default probability. Finally note that the slope of the curve has to be adjusted for the industrial sector.

6.3 Dynamics of migration probabilities

In Section 6.2 we have analyzed the sequence of migration probabilities under the assumption of i.i.d. transition matrices. This assumption, which is usually adopted for computing migration correlations (see Section 4), can be questioned in practice. The aim of this section is to highlight the dynamics of migration probabilities before estimating a dynamic factor model in Section 6.4.

In the first subsection we plot the series of up- and down-grade probabilities, and discuss their serial dependence. Then in Section 6.3.2 we consider their relationship with the French business cycle.

6.3.1 The evolution of up- and downgrade probabilities

Let us focus on downgrade and upgrade probabilities involving a migration of at most 2 buckets: $d_{k,t} = \pi_{k,k-1,t} + \pi_{k,k-2,t}$, $u_{k,t} = \pi_{k,k+1,t} + \pi_{k,k+2,t}$, respectively [except for the extreme classes, where the number of buckets is one]. The time series of downgrade and upgrade probabilities are reported in Figures 3 and 4, respectively.

[Insert Figure 3: Downgrade probabilities]

[Insert Figure 4: Upgrade probabilities]

Each panel corresponds to a rating class at the beginning of the year and provides the dynamics in the wholesale (circles) and retail trade (diamonds) sectors. The first and second order autocorrelations are reported in Tables 11 and 12 for the different downgrade and upgrade series and for the wholesale and retail trade sectors, respectively:

[Insert Table 11: Autocorrelations in the wholesale sector]

[Insert Table 12: Autocorrelations in the retail trade sector]

Despite the rather small number of observation dates, it is immediately seen that the serial autocorrelations are rather high. Thus the usual independence assumption, which underlies the computation of migration correlations, does not seem to be relevant. However the source of serial dependence can have a dimension which is much smaller than the number of migration probabilities reported in Figures 3 and 4. A first estimate of the number of underlying dynamic factors can be derived by considering a singular value decomposition of the historical variance-covariance matrix of the transition probabilities per industrial sector. The eigenvalues ranked by decreasing order are reported in Table 13 and 14.

[Insert Table 13: Singular value decomposition in the wholesale sector]

[Insert Table 14: Singular value decomposition in the retail trade sector]

The first eigenvalue is much larger than the next ones, which indicates a dominant factor. The second eigenvalue is also significant and the next ones can be neglected in a first step. Moreover the levels of the eigenvalues are similar in the two sectors. Thus a one- or two-factor model is expected[29].

Such a factor analysis could have been improved by considering jointly two sectors for instance and separating the common factors from the idiosyncratic ones. This would allow not only to account for the common movements within each sector as considered in the paper, but also for common movements between sectors. This more complete analysis is out of the scope of the paper.

[29] This singular value decomposition will be improved in Section 6.4 by applying a preliminary nonlinear transformation to all series.

6.3.2 Business cycle

It is usual to relate the failure rate with the general state of the economy, that is the so-called business cycles. The Banque de France data base on migration probabilities convey much more information and a better knowledge of the link with business cycles can be expected. Two recent studies have already been performed on the data by Standard & Poor's, and Moody's, respectively, with proxies of the US business cycle [see Bangia et alii (2002), and Nickell, Perraudin, Variotto (2000), respectively]. Even if this relationship is not the topic of our paper, it is interesting to give some preliminary elements. In a first step the dynamics of downgrade and upgrade probabilities can be compared with the evolution of GDP in France in the period 1992-2002, which is provided in Figure 5.

[Insert Figure 5: GDP evolution in France]

The dynamic linear link between the series can be studied by a causality analysis between the downgrade (resp. upgrade) series and the GDP. The lead and lagged causality measures[30] of downgrade and upgrade probabilities with GDP increment I_t are reported in Tables 15 and 16 for the wholesale and the retail trade sector, respectively.

[Insert Table 15: Causality relations in the wholesale sector]

[Insert Table 16: Causality relations in the retail trade sector]

It is seen that the distribution of causality measures is different for up- and downgrades, and for the different classes of risk. The downgrades are generally more reactive than the upgrades. Moreover for the classes with low risk the causality from I to d, say, is more important than the causality from d to I. For instance the business cycle affects the downgrades for class 7. But the ordering between both directional measures is reversed in the very risky classes, where the downgrade probabilities provide a leading indicator of the business cycle with a lead between 2 and 3 years.

[30] See Gourieroux, Jasiak (2001) for a description of causality including lead and lagged measures.

6.4 Estimation of the ordered probit model with unobservable factor

Finally we estimate the ordered probit model with unobservable factor introduced in Section 3.2 and more specifically its approximation for large cross-sectional dimension presented in Section 5.4. The estimation is performed for the wholesale sector.

In a first step we compute the transformed series $y_{kl,t} = G^{-1}(\widehat{\pi}^*_{kl,t}), \forall k, l$, and perform their principal component analysis, that is the spectral decomposition of matrix $\widetilde{Y}\widetilde{Y}'$, where the rows of $\widetilde{Y}$ are given by $y_{kl,t} - \overline{y}_{kl}$, k, l varying, with $\overline{y}_{kl} = \frac{1}{T}\sum_t y_{kl,t}$. The corresponding eigenvalues are given in decreasing order in the following table:

5.963	2.740	2.166	0.739	0.393	0.314	0.204	0.125	0.051	0

Three eigenvalues are much larger than the other ones[31]. The normalized eigenvectors corresponding to the 3 largest eigenvalues are displayed in Figure 6.

[Insert Figure 6: Factors corresponding to the 3 largest eigenvalues]

The pattern of the factor corresponding to the largest eigenvalue is consistent with the evolution of downgrade probabilities reported in Figure 3, for all rating classes except the riskiest one (class 1). Indeed the factor points out an overall decreasing downgrade risk over the sample period, with peaks of downgrade probabilities in 1994 and in 1998-1999 [32]. The factor corresponding to the second eigenvalue feature a similar pattern, but the peaks occur in 1995-1996 and 1999-2000. In particular the peak in 1995-1996 may be associated with the large downgrade probabilities featured by class 1 (the riskiest class) in those years (see Figure 3). Finally it is important to see how the "business cycle" is related to the three factors. For this purpose the relative change in GDP has been regressed on the constant and the three

[31] Note that a principal component analysis (PCA) applied to series π or to transformed series $G^{-1}(\pi)$ does not necessarily provide the same number of factors and of course the same factors. This is due to the linear factorial interpretation of PCA.

[32] The sign of the factor has been chosen so that the corresponding estimated coefficients β_k are positive for each class. Thus the larger is the factor value, the larger are the downgrade probabilities.

factors. The regression coefficients are:

$$I_t = 1.970 - 0.303Z_{1,t} - 0.481Z_{2,t} + 0.046Z_{3,t},$$

with $R^2 = 0.190$. This regression analysis and the comparison with the pattern of GDP increments displayed in Figure 5 suggest that the factors corresponding to the two largest eigenvalues are related to the business cycle. Indeed the overall improvement in credit quality in 1992-2001 suggested by the factors is associated with the positive trend in GDP increments over the same period. Moreover the peak in downgrade risk in 1995-1996 may be related to the slow down in GDP increment in these years. However the evolutions of credit cycle and business cycle are not fully parallel [see Feng et alii (2003) for similar findings in US data]. For instance the peak in downgrade risk in 1998-1999 anticipates the slow growth years 2001-2002. This explains the rather poor fit in the regression.

The analysis is completed by applying the Kalman filter with 3 factors. This provides the dynamics of the factors and the estimated structural parameters, which are reported in the following table:

Table 17: Structural parameters

$c_2 = 0$	$c_3 = 0.6611$	$c_4 = 0.9470$	$c_5 = 1.1776$	$c_6 = 1.3714$
$c_7 = 1.4680$	$c_8 = 1.5662$			

$\alpha_1 = -0.4410$	$\beta_{1,1} = 0.0210$	$\beta_{2,1} = 0.0389$	$\beta_{3,1} = 0.0097$	$\sigma_1 = 1$
$\alpha_2 = 0.6843$	$\beta_{1,2} = 0.0020$	$\beta_{2,2} = 0.0091$	$\beta_{3,2} = 0.0065$	$\sigma_2 = 0.4067$
$\alpha_3 = 1.0026$	$\beta_{1,3} = 0.0086$	$\beta_{2,3} = 0.0066$	$\beta_{3,3} = 0.0049$	$\sigma_3 = 0.2621$
$\alpha_4 = 1.1589$	$\beta_{1,4} = 0.0066$	$\beta_{2,4} = 0.0073$	$\beta_{3,4} = 0.0048$	$\sigma_4 = 0.2316$
$\alpha_5 = 1.2902$	$\beta_{1,5} = 0.0280$	$\beta_{2,5} = -0.0010$	$\beta_{3,5} = 0.0033$	$\sigma_5 = 0.2039$
$\alpha_6 = 1.3884$	$\beta_{1,6} = 0.0252$	$\beta_{2,6} = 0.0050$	$\beta_{3,6} = 0.0002$	$\sigma_6 = 0.1655$
$\alpha_7 = 1.4766$	$\beta_{1,7} = 0.0134$	$\beta_{2,7} = 0.0025$	$\beta_{3,7} = 0.0157$	$\sigma_7 = 0.1712$

$A_{11} = 0.2538$	$A_{12} = -0.0087$	$A_{13} = -0.1887$
$A_{21} = -0.7646$	$A_{22} = -0.0832$	$A_{23} = -0.2578$
$A_{31} = -0.0675$	$A_{32} = -0.1186$	$A_{33} = 0.0324$

As expected the estimated thresholds c are increasing. Similarly the intercepts are increasing with respect to the rating index k, which confirms that downgrade risk is higher for the lower rating classes. It is seen that the β_1 coefficients are higher and more homogeneous, and thus the first factor appears as a general factor. The β_2 coefficients show some opposition between the classes 1 - 4 and the classes 5 - 7, that is between speculative and investment categories. Finally the volatility parameters σ are generally smaller for the riskier rating classes.

Compared to the standard use of the ordered probit model suggested by the Basle Committee, we have followed an approach which is in favour of this model:

i) 3 factors have been introduced instead of a single factor as usual;

ii) the factors have been defined endogenously by a principal component analysis and not chosen a priori,

iii) the estimation has been performed on a given industrial sector, likely more homogenous than the whole population.

Nevertheless it is seen in Figures 7-13, which provide the actual and fitted migration probabilities, that the fit is not entirely satisfactory (note that the scales are not the same on the different figures).

[Insert Figures 7-13: Actual and fitted cumulated migration probabilities]

It does not mean that the ordered polytomous model has to be rejected, but there is still some specification errors. Some possible ones are the following:

i) the population is not sufficiently homogenous,

ii) the factors Z_t can have an instantaneous effect by means of Z_t, but also lagged ones by means of $Z_{t-1}, Z_{t-2}, ...$ Such lagged effects likely exist following the causality analysis of Section 6.3,

iii) the latent distribution G can be different from the Gaussian one, and could feature different tail or skewness behaviours as in the standard affine stochastic intensity models (see Section 3.2),

iv) the latent distribution G can depend on the starting rating class. This is likely the main specification error as seen in Figures 8-10, for rating classes 2, 3, 4 for instance. Indeed the general patterns of the actual migration probabilities are almost satisfactory, but some of them differ by a drift. This drift can be corrected by means of an appropriate choice of the G function.

These various specification tests will be left for further research.

7 Concluding remarks

The stochastic migration model introduced in this paper is a specification which is flexible and appropriate for the joint analysis of rating migration of several firms. We have discussed several properties of the MST model concerning the prediction of rating transitions, the migration correlations, the links between latent and observable correlations, and some special features on statistical inference. As an illustration two models have been estimated on the French data set of the Banque de France. We have essentially considered the two type of models suggested in the literature. The model with i.i.d. stochastic transition matrices underlies the computation of migration correlations, but is clearly misspecified, due in particular, but not only, to the effect of the business cycle, which induces serial dependence. We have also performed one of the first estimations of the ordered probit model with unobservable factors and given some direction of future research for improving this approach.

Finally note that the MST model is not only useful for prediction purpose, but can be used in a more structural way. For instance, in credit risk, it can be the basis for pricing and hedging strategies of credit portfolios and credit derivatives [see e.g. Jarrow, Lando, Turnbull (1997), Del Angel, Diez-Canedo, Gorbea (1998)].

REFERENCES

Albanese, C., Campolieti, J., Chen, O., and A., Zavidonov (2003): "Credit Barrier Models", University of Toronto DP.

Amemiya, T. (1976): "The Maximum Likelihood, the Minimum Chi-Square, and the Nonlinear Weighted Least Squares Estimators in the General Qualitative Response Model", JASA, 71, 347-351.

Anderson, T., and L., Goodman (1957): "Statistical Inference about Markov Chains", Annals of Math. Statistics, 28, 89-110.

Bahar, R., and K., Nagpal (2001): "Measuring Default Correlation", Risk, March, 129-132.

Bangia, A., Diebold, F., Kronimus, A., Schlagen, C., and T., Schuerman (2002): "Rating Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing", Journal of Banking and Finance, 26, 445-474.

Bardos, M., Foulcher, S., and E., Bataille (2004): "Les Scores de la Banque de France: Méthode, Résultats, Applications", Banque de France, Observatoire des Entreprises.

Basle Committee on Banking Supervision (2002), *Quantitative Impact Study 3. Technical Guidance*, Banking for International Settlements.

Berkson, J. (1944): "Application of the Logistic Function to Bio-assay", JASA, 39, 357-365.

Brady, B., and R., Bos (2002): "Record Default in 2001: the Result of Poor Credit Quality and a Weak Economy", Standard and Poor's, February.

Crouhy, M., Galai, D., and R. Mark (2000): "A Comparative Analysis of Current Credit Risk Models", Journal of Banking and Finance, 29, 59-117.

Del Angel, G., Diez-Canedo, J., and E., Gorbea (1998): "A Discrete Markov Chain Model for Valuing Loan Portfolios", Journal of Banking and Finance, 22, 1457-1480.

De Servigny, A., and O., Renault (2002): "Default Correlation: Empirical Evidence", Standard and Poor's, September.

Duffie, D., and D., Lando (2001): "Term Structure of Credit Spreads with Incomplete Accounting Information", Econometrica, 69, 633-664.

Duffie, D., and K., Singleton (1999): "Modeling the Term Structure of Defaultable Bonds", The Review of Financial Studies, 12, 687-720.

Feng, D., Gouriéroux, C., and J., Jasiak (2003): "The Ordered Qualitative Model for Credit Rating Transitions", CREST DP.

Foulcher, S., Gouriéroux, C., and A., Tiomo (2003): "The Term Structure of Defaults and Ratings", Banque de France, forthcoming Insurance and Risk Management.

Frey, R., and A., McNeil (2001): "Modelling Dependent Defaults", Working Paper, ETH Zurich.

Frey, A., and A., McNeil (2003): "Dependent Defaults in Models of Portfolio Credit Risk", forthcoming in Journal of Risk.

Gagliardini, P., and C., Gouriéroux (2003): "Migration Correlation: Definition and Consistent Estimation", forthcoming Journal of Banking and Finance.

Gordy, J., and E., Heitfield (2001): "Of Moody's and Merton: a Structural Model of Bond Rating Transitions", Working Paper, Federal Reserve.

Gordy, J., and E., Heitfield (2002): "Estimating Default Correlation from Short Panels of Credit Rating Performance Data", Working Paper, Federal Reserve Board.

Gouriéroux, C., and J., Jasiak (2001): *Financial Econometrics,* Princeton University Press.

Gouriéroux, C., and J., Jasiak (2003a): *Microeconometrics for Finance, Insurance, and Marketing*, forthcoming Princeton University Press.

Gouriéroux, C., and J., Jasiak (2003b): "Jacobi Process with Application to Switching Regimes", forthcoming Journal of Econometrics.

Gouriéroux, C., and A., Monfort (1995): *Simulation Based Econometric Methods,* Louvain, CORE Lecture Series, Oxford University Press.

Gouriéroux, C., and A., Monfort (2002): "Equidependence in Qualitative and Duration Models with Application to Credit Risk", CREST DP.

Gouriéroux, C., and A., Monfort (2004): "Error-in-Factor Models", CREST DP.

Gouriéroux, C., Monfort, A., and V., Polimenis (2003): "Affine Models for Credit Risk", CREST DP.

Gupton, G., Finger, C., and M., Bahtia (1997): "Creditmetrics-technical document", Technical Report, the RiskMetrics Group.

Jarrow, R., Lando, D., and S., Turnbull (1997): "A Markov Model for the Term Structure of Credit Risk Spread", Review of Financial Studies, 10, 481-523.

Joe, H. (1997): *Multivariate Models and Dependence Concepts,* Monographs on Statistics and Applied Probability, 73, Chapman & Hall.

Lando, D. (1998): "On Cox Processes and Credit Risky Securities", Review of Derivative Research, 2, 99-120.

Lucas, D. (1995a): "Default Correlation and Credit Analysis", Journal of Fixed Income, March, 76-87.

Lucas, D. (1995b): "The Effectiveness of Downgrade Provisions in Reducing Counterparty Risk", Journal of Fixed Income, June, 32-41.

McRae, E. (1977): "Estimation of Time Varying Markov Processes with Aggregated Data", Econometrica, 45, 183-198.

Merton, R. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", Journal of Finance, 20, 449-470.

Nickell, P., Perraudin, W., and S. Variotto (2000): "Stability of Rating Transitions", Journal of Banking and Finance, 24, 203-227.

Schonbucher, P. (2000): "Factor Models for Portfolio Credit Risk", University of Bonn DP.

Appendix 1
Migration correlation in the beta model

In this Appendix we derive the up-up, down-down, up-down migration correlation matrices, as well as the migration correlations with identical initial and final ratings, at horizon 1 in the beta model.

a) Up-up migration correlation matrix

From (6), (12), and the independence of the rows in the beta model we get:

$$\begin{aligned}
& corr\left(\mathbb{I}_{Y_{i,t+1}=k+1}, \mathbb{I}_{Y_{j,t+1}=l+1} \mid Y_{i,t}=k, Y_{j,t}=l\right) \\
= \; & cov\left(\pi_{k,k+1,t}, \pi_{l,l+1,t}\right) \frac{1}{\left[E\pi_{k,k+1,t}\left(1-E\pi_{k,k+1,t}\right)\right]^{1/2}} \frac{1}{\left[E\pi_{l,l+1,t}\left(1-E\pi_{l,l+1,t}\right)\right]^{1/2}} \\
= \; & \delta_{kl} \frac{\alpha_{k,k+1}\left(1-\alpha_{k,k+1}\right)}{1+\gamma_k} \frac{1}{\left[\alpha_{k,k+1}\left(1-\alpha_{k,k+1}\right)\right]^{1/2}} \frac{1}{\left[\alpha_{k,k+1}\left(1-\alpha_{k,k+1}\right)\right]^{1/2}} \\
= \; & \delta_{kl} \frac{1}{1+\gamma_k}.
\end{aligned}$$

b) Down-down migration correlation matrix

Similarly we get:

$$\begin{aligned}
& corr\left(\mathbb{I}_{Y_{i,t+1}=k-1}, \mathbb{I}_{Y_{j,t+1}=l-1} \mid Y_{i,t}=k, Y_{j,t}=l\right) \\
= \; & cov\left(\pi_{k,k-1,t}, \pi_{l,l-1,t}\right) \frac{1}{\left[E\pi_{k,k-1,t}\left(1-E\pi_{k,k-1,t}\right)\right]^{1/2}} \frac{1}{\left[E\pi_{l,l-1,t}\left(1-E\pi_{l,l-1,t}\right)\right]^{1/2}} \\
= \; & \delta_{kl} \frac{\alpha_{k,k-1}\left(1-\alpha_{k,k-1}\right)}{1+\gamma_k} \frac{1}{\left[\alpha_{k,k-1}\left(1-\alpha_{k,k-1}\right)\right]^{1/2}} \frac{1}{\left[\alpha_{k,k-1}\left(1-\alpha_{k,k-1}\right)\right]^{1/2}} \\
= \; & \delta_{kl} \frac{1}{1+\gamma_k}.
\end{aligned}$$

c) Up-down migration correlation matrix

We have:

$$
\begin{aligned}
& corr\left(\mathbb{I}_{Y_{i,t+1}=k+1}, \mathbb{I}_{Y_{j,t+1}=l-1} \mid Y_{i,t}=k, Y_{j,t}=l\right) \\
= \; & cov\left(\pi_{k,k+1,t}, \pi_{l,l-1,t}\right) \frac{1}{\left[E\pi_{k,k+1,t}\left(1-E\pi_{k,k+1,t}\right)\right]^{1/2}} \frac{1}{\left[E\pi_{l,l-1,t}\left(1-E\pi_{l,l-1,t}\right)\right]^{1/2}} \\
= \; & -\delta_{kl} \frac{\alpha_{k,k+1}\alpha_{k,k-1}}{1+\gamma_k} \frac{1}{\left[\alpha_{k,k+1}\left(1-\alpha_{k,k+1}\right)\right]^{1/2}} \frac{1}{\left[\alpha_{k,k-1}\left(1-\alpha_{k,k-1}\right)\right]^{1/2}} \\
= \; & \delta_{kl} \frac{1}{1+\gamma_k} \left[\frac{\alpha_{k,k+1}}{1-\alpha_{k,k+1}}\right]^{1/2} \left[\frac{\alpha_{k,k-1}}{1-\alpha_{k,k-1}}\right]^{1/2}.
\end{aligned}
$$

d) Migration correlation with identical starting and final ratings

Finally from (13) we get:

$$
\begin{aligned}
& corr\left(\mathbb{I}_{Y_{i,t+1}=k^*}, \mathbb{I}_{Y_{j,t+1}=k^*} \mid Y_{i,t}=k, Y_{j,t}=k\right) \\
= \; & corr\left(\pi_{kk^*,t}, \pi_{kk^*,t}\right) \left[\frac{V\pi_{kk^*,t}}{E\pi_{kk^*,t}\left(1-E\pi_{kk^*,t}\right)}\right]^{1/2} \left[\frac{V\pi_{kk^*,t}}{E\pi_{kk^*,t}\left(1-E\pi_{kk^*,t}\right)}\right]^{1/2} \\
= \; & \frac{1}{1+\gamma_k}.
\end{aligned}
$$

Appendix 2
The count process

In this Appendix we consider the count process $(N_{k,t}, k = 1, ..., K)$.

i) Markov property

Let us first prove that $(N_{k,t}, k = 1, ..., K)$ is a Markov process. Let us consider the count variables $N_{kl,t+1}$, $k, l = 1, ..., K$. Their distribution conditional to the lagged ratings and to the transition matrices is given by:

$$\begin{aligned} & l\left(N_{kl,t+1}, k, l = 1, ..., K \mid \underline{Y_t}, (\Pi_t)\right) \\ = & \prod_{k=1}^{K}\left(\frac{N_{k,t}!}{\prod_{l=1}^{K} N_{kl,t+1}!}\prod_{l=1}^{K}\pi_{kl,t+1}^{N_{kl,t+1}}\right), \end{aligned}$$

that is the product of K multinomial distributions of orders $N_{k,t}$ and parameters $(\pi_{kl,t+1}, l = 1, ..., K)$, $k = 1, ..., K$.

The distribution of $N_{kl,t+1}$, $k, l = 1, ..., K$ given the lagged ratings is obtained by integrating out the transition matrix:

$$\begin{aligned} & l\left(N_{kl,t+1}, k, l = 1, ..., K \mid \underline{Y_t}\right) \\ = & \prod_{k=1}^{K}\left(\frac{N_{k,t}!}{\prod_{l=1}^{K} N_{kl,t+1}!}E\left[\prod_{l=1}^{K}\pi_{kl,t+1}^{N_{kl,t+1}}\right]\right). \end{aligned}$$

In particular, it depends on the past through the count variables $N_{k,t}$, $k = 1, ..., K$, only. Since $N_{l,t+1} = \sum_{k=1}^{K} N_{kl,t+1}$, we deduce that $(N_{1,t}, ..., N_{K,t})$, t varying, is a Markov process.

ii) First conditional moment

Let us derive the first conditional moment of the Markov process $(N_{1,t}, ..., N_{K,t})$. It is given by:

$$\begin{aligned} \underset{t}{E}\left[N_{l,t+1}\right] &= E\left(E\left[N_{l,t+1} \mid N_t, (\Pi_t)\right] \mid N_t\right) = \sum_{k=1}^{K} N_{k,t} E\left(E\left[\pi_{kl,t+1} \mid N_t, (\Pi_t)\right] \mid N_t\right) \\ &= \sum_{k=1}^{K} N_{k,t} E\pi_{kl,t}, \end{aligned}$$

where $N_t = (N_{1,t}, ..., N_{K,t})'$.

Table 1
1-year transition matrix

	7	6	5	4	3	2	1	0	NR
7	0.7155	0.0979	0.0241	0.0113	0.0031	0.0002	0.0002	0.0002	0.1474
6	0.1226	0.5977	0.1031	0.0321	0.0187	0.0022	0.0008	0.0011	0.1216
5	0.0169	0.2544	0.4232	0.1145	0.0477	0.0079	0.0025	0.0026	0.1305
4	0.0085	0.0600	0.2519	0.3517	0.1223	0.0297	0.0109	0.0070	0.1579
3	0.0011	0.0413	0.0924	0.2450	0.2934	0.0755	0.0339	0.0200	0.1974
2	0.0000	0.0114	0.0509	0.1500	0.2500	0.1640	0.0842	0.0307	0.2588
1	0.0000	0.0076	0.0317	0.0544	0.1903	0.1224	0.1813	0.0650	0.3474
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000

Table 1: 1-year transition matrix for the wholesale sector in year 2001.

Table 2
Adjusted 1-year transition matrix

	7	6	5	4	3	2	1	0
7	0.8392	0.1148	0.0282	0.0133	0.0036	0.0003	0.0003	0.0003
6	0.1396	0.6804	0.1174	0.0366	0.0213	0.0025	0.0009	0.0013
5	0.0194	0.2925	0.4867	0.1316	0.0549	0.0090	0.0029	0.0030
4	0.0101	0.0713	0.2991	0.4177	0.1452	0.0352	0.0130	0.0084
3	0.0014	0.0514	0.1152	0.3053	0.3656	0.0940	0.0422	0.0249
2	0.0000	0.0154	0.0686	0.2024	0.3373	0.2213	0.1136	0.0414
1	0.0000	0.0116	0.0486	0.0833	0.2917	0.1875	0.2778	0.0995
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 2: Adjusted 1-year transition probabilities for the wholesale sector in 2001.

Table 3
Average 1-year transition matrix

	7	6	5	4	3	2	1	0
7	0.7968	0.1491	0.0317	0.0165	0.0040	0.0007	0.0003	0.0009
6	0.1227	0.6476	0.1458	0.0515	0.0272	0.0030	0.0010	0.0012
5	0.0162	0.2388	0.4935	0.1627	0.0703	0.0125	0.0032	0.0028
4	0.0084	0.0597	0.2587	0.4371	0.1795	0.0373	0.0124	0.0069
3	0.0026	0.0344	0.1002	0.2777	0.4095	0.1136	0.0430	0.0190
2	0.0017	0.0129	0.0494	0.1642	0.3420	0.2524	0.1338	0.0436
1	0.0015	0.0102	0.0276	0.0937	0.2432	0.2037	0.3312	0.0889
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 3: Estimated average 1-year transition matrix for the wholesale sector.

Table 4
Average 1-year transition matrix

	7	6	5	4	3	2	1	0
7	0.7385	0.2175	0.0299	0.0071	0.0054	0.0007	0.0004	0.0005
6	0.0767	0.7301	0.1353	0.0263	0.0267	0.0029	0.0009	0.0010
5	0.0086	0.2796	0.4893	0.1025	0.0963	0.0154	0.0059	0.0024
4	0.0057	0.1081	0.3556	0.2267	0.2388	0.0403	0.0188	0.0060
3	0.0026	0.0669	0.1988	0.1704	0.3942	0.1002	0.0514	0.0155
2	0.0009	0.0285	0.1040	0.1111	0.3466	0.2051	0.1659	0.0379
1	0.0008	0.0141	0.0563	0.0576	0.2363	0.1842	0.3647	0.0861
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 4: Estimated average 1-year transition matrix for the retail trade sector.

Table 5
Joint upgrade probabilities

	7	6	5	4	3	2	1	0
7	–	–	–	–	–	–	–	–
6	–	0.0153	0.0295	0.0318	0.0341	0.0420	0.0249	–
5	–	0.0295	0.0574	0.0621	0.0665	0.0816	0.0485	–
4	–	0.0318	0.0621	0.0673	0.0721	0.0885	0.0525	–
3	–	0.0341	0.0665	0.0721	0.0775	0.0950	0.0565	–
2	–	0.0420	0.0816	0.0885	0.0950	0.1171	0.0697	–
1	–	0.0249	0.0485	0.0525	0.0565	0.0697	0.0418	–
0	–	–	–	–	–	–	–	–

Table 5: Joint upgrade probabilities for two firms in the wholesale sector. Row and column numbers denote the initial rating class of the two firms.

Table 6
Upgrade correlations

	7	6	5	4	3	2	1	0
7	–	–	–	–	–	–	–	–
6	–	0.0018	0.0013	0.0005	0.0000	0.0000	−0.0006	–
5	–	0.0013	0.0023	0.0017	0.0012	−0.0001	−0.0010	–
4	–	0.0005	0.0017	0.0017	0.0014	0.0001	−0.0010	–
3	–	0.0000	0.0012	0.0014	0.0016	0.0002	−0.0006	–
2	–	0.0000	−0.0001	0.0001	0.0002	0.0008	0.0004	–
1	–	−0.0006	−0.0010	−0.0010	−0.0006	0.0004	0.0020	–
0	–	–	–	–	–	–	–	–

Table 6: Upgrade correlations for two firms in the wholesale sector. The row and column numbers denote the initial rating classes of the two firms.

Table 7
Joint downgrade probabilities

	7	6	5	4	3	2	1	0
7	0.0225	0.0219	0.0244	0.0269	0.0170	0.0199	0.0132	–
6	0.0219	0.0214	0.0239	0.0263	0.0167	0.0196	0.0130	–
5	0.0244	0.0239	0.0267	0.0295	0.0187	0.0220	0.0145	–
4	0.0269	0.0263	0.0295	0.0326	0.0207	0.0243	0.0161	–
3	0.0170	0.0167	0.0187	0.0207	0.0131	0.0155	0.0102	–
2	0.0199	0.0196	0.0220	0.0243	0.0155	0.0183	0.0120	–
1	0.0132	0.0130	0.0145	0.0161	0.0102	0.0120	0.0081	–
0	–	–	–	–	–	–	–	–

Table 7: Joint downgrade probabilities for two firms in the wholesale sector. Row and column numbers denote the initial rating class of the two firms.

Table 8
Downgrade correlations

	7	6	5	4	3	2	1	0
7	0.0021	0.0013	0.0011	0.0008	0.0006	0.0000	−0.0004	–
6	0.0013	0.0011	0.0010	0.0011	0.0009	0.0007	0.0003	–
5	0.0011	0.0010	0.0016	0.0019	0.0017	0.0016	0.0003	–
4	0.0008	0.0011	0.0019	0.0027	0.0024	0.0026	0.0010	–
3	0.0006	0.0009	0.0017	0.0024	0.0024	0.0025	0.0011	–
2	0.0000	0.0007	0.0016	0.0026	0.0025	0.0034	0.0011	–
1	−0.004	0.0003	0.0003	0.0010	0.0011	0.0011	0.0031	–
0	–	–	–	–	–	–	–	–

Table 8: Downgrade correlations for two firms in the wholesale sector. The row and column numbers denote the initial rating classes of the two firms.

Table 9
Default correlation in the wholesale sector

	7	6	5	4	3	2	1	0
7	0.0004	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000	–
6	0.0002	0.0004	0.0002	0.0002	0.0002	0.0000	0.0000	–
5	0.0002	0.0002	0.0003	0.0001	0.0002	−0.0001	0.0003	–
4	0.0000	0.0002	0.0001	0.0003	0.0002	0.0002	0.0005	–
3	0.0000	0.0002	0.0002	0.0002	0.0006	−0.0003	0.0009	–
2	0.0000	0.0000	−0.0001	0.0002	−0.0003	0.0006	0.0000	–
1	0.0000	0.0000	0.0003	0.0005	0.0009	0.0000	0.0031	–
0	–	–	–	–	–	–	–	–

Table 9: Default correlations for two firms in the wholesale sector. The row and column numbers denote the initial rating classes of the two firms.

Table 10
Default correlation in the retail trade sector

	7	6	5	4	3	2	1	0
7	0.0009	0.0000	0.0003	0.0003	0.0009	0.0007	0.0005	–
6	0.0000	0.0001	0.0000	0.0001	0.0000	0.0000	0.0001	–
5	0.0003	0.0000	0.0003	0.0000	0.0004	0.0003	0.0003	–
4	0.0003	0.0001	0.0000	0.0007	0.0006	0.0004	0.0006	–
3	0.0009	0.0000	0.0004	0.0006	0.0013	0.0010	0.0011	–
2	0.0007	0.0000	0.0003	0.0004	0.0010	0.0014	0.0007	–
1	0.0005	0.0001	0.0003	0.0006	0.0011	0.0007	0.0014	–
0	–	–	–	–	–	–	–	–

Table 10: Default correlations for two firms in the retail trade sector. The row and column numbers denote the initial rating classes of the two firms.

Table 11
Autocorrelations in the wholesale sector

		7	6	5	4	3	2	1
down	order 1	0.0430	0.7902	0.6907	0.4373	0.2541	-0.2316	0.2516
	order 2	0.3182	0.3571	0.6878	0.5267	0.4373	0.1686	-0.3434
up	order 1	-	0.0551	0.3146	0.3815	0.2800	0.0283	0.2059
	order 2	-	-0.2284	0.0627	0.1954	0.4626	0.5549	0.0974

Table 11: Autocorrelations of order 1 and 2 for downgrade and upgrade probabilities in the wholesale sector.

Table 12
Autocorrelations in the retail trade sector

		7	6	5	4	3	2	1
down	order 1	0.0521	0.7573	0.4433	0.2096	0.4497	0.7131	-0.2307
	order 2	-0.1677	0.6729	0.3902	0.3687	0.5290	0.3226	-0.1481
up	order 1	-	0.4506	0.0959	0.0730	0.4863	0.1014	0.6707
	order 2	-	0.3620	0.3880	0.3867	0.4573	0.2222	0.3436

Table 12: Autocorrelations of order 1 and 2 for downgrade and upgrade probabilities in the retail trade sector.

Table 13
Singular value decomposition in the wholesale sector

0.00544	0.00101	0.00037	0.00025	0.00011	0.00007	0.00005
0.00001	2E-6	1E-19	1E-20	<1E-20	<1E-20	

Table 13: Eigenvalues of the variance-covariance matrix of the series of down- and upgrade probabilities in the wholesale sector.

Table 14
Singular value decomposition in the retail trade sector

0.00719	0.00071	0.00036	0.00012	0.00009	0.00007	0.00004
0.00002	7E-6	-4E-19	<1E-20	<1E-20	<1E-20	

Table 14: Eigenvalues of the variance-covariance matrix of the series of down- and upgrade probabilities in the retail trade sector.

Table 15
Causality relations in the wholesale sector

	7	6	5	4	3	2	1
$C_{I \to d}$	**10.51**	**5.62**	**5.79**	**5.79**	**7.27**	**4.68**	0.47
$C_{d \to I}(1)$	1.18	0.44	0.08	0.47	1.47	3.25*	1.02
$C_{d \to I}(2)$	1.54	1.86	0.02	0.14	0.52	2.43	3.10*
$C_{d \to I}(3)$	2.39	**5.12**	2.74*	3.51*	3.07*	2.12	**5.79**
$C_{I \to u}$	–	2.27	**8.97**	**6.72**	**7.96**	**7.69**	0.44
$C_{u \to I}$	–	0.34	1.38	1.38	1.52	2.36	0.36

Table 15: Causality relations in the wholesale sector. The causality measures are multiplied by T. Under the null hypothesis of no linear link, these standardized statistics are asymptotically $\chi^2(1)$-distributed. Bold entries (resp. entries with an asterisk) correspond to significant values at the 5% (10%, respectively) level. For the causality directions $d \to I$ we provide measures at horizon $h = 1, 2, 3$.

Table 16
Causality relations in the retail trade sector

	7	6	5	4	3	2	1
$C_{I \to d}$	**29.28**	2.74	2.70	**4.08**	2.11	0.24	0.01
$C_{d \to I}(1)$	1.75	0.11	0.20	0.81	0.01	1.50	0.98
$C_{d \to I}(2)$	1.82	0.23	0.17	0.26	0.00	0.80	**6.54**
$C_{d \to I}(3)$	2.58	1.06	2.49	1.97	2.51	2.44	2.06
$C_{I \to u}$	–	**8.01**	**4.43**	3.42	1.39	0.82	0.42
$C_{u \to I}$	–	1.58	1.18	1.23	0.00	0.62	0.23

Table 16: Causality relations in the retail trade sector. The causality measures are multiplied by T. Under the null hypothesis of no linear link, these standardized statistics are asymptotically $\chi^2(1)$-distributed. Bold entries correspond to significant values at the 5% level. For the causality directions $d \to I$ we provide measures at horizon $h = 1, 2, 3$.

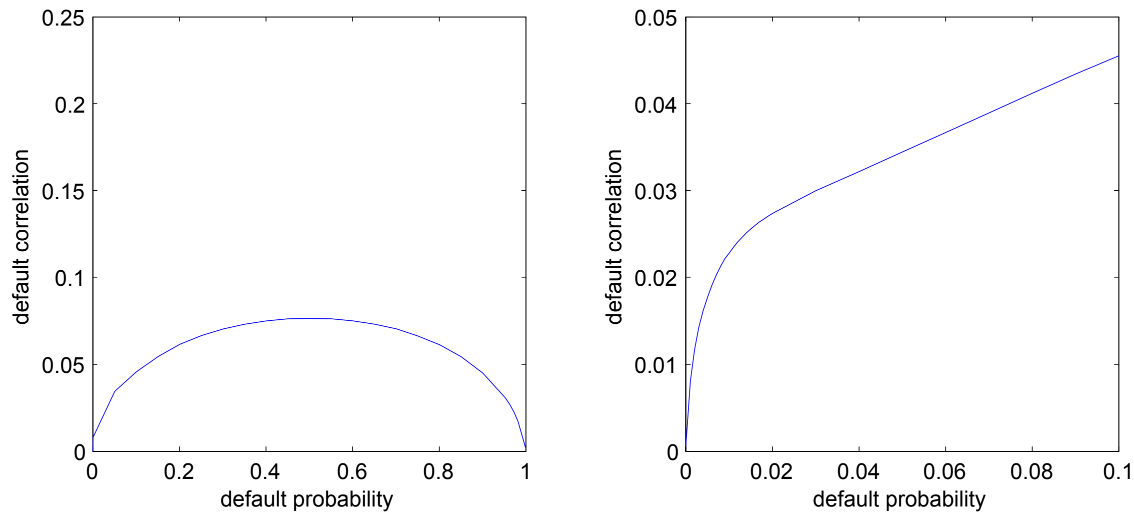

Figure 1: Default correlation as a function of default probability by Basle Committee.

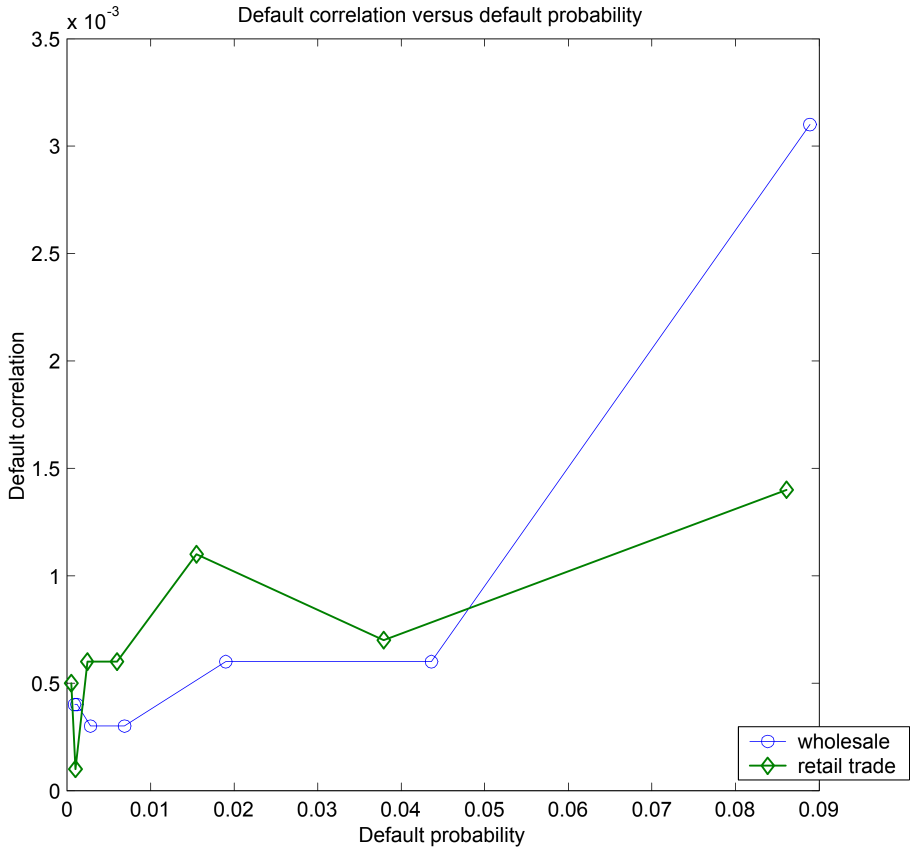

Figure 2: Default correlation as a function of default probability across the different rating classes for the wholesale and the retail trade sectors.

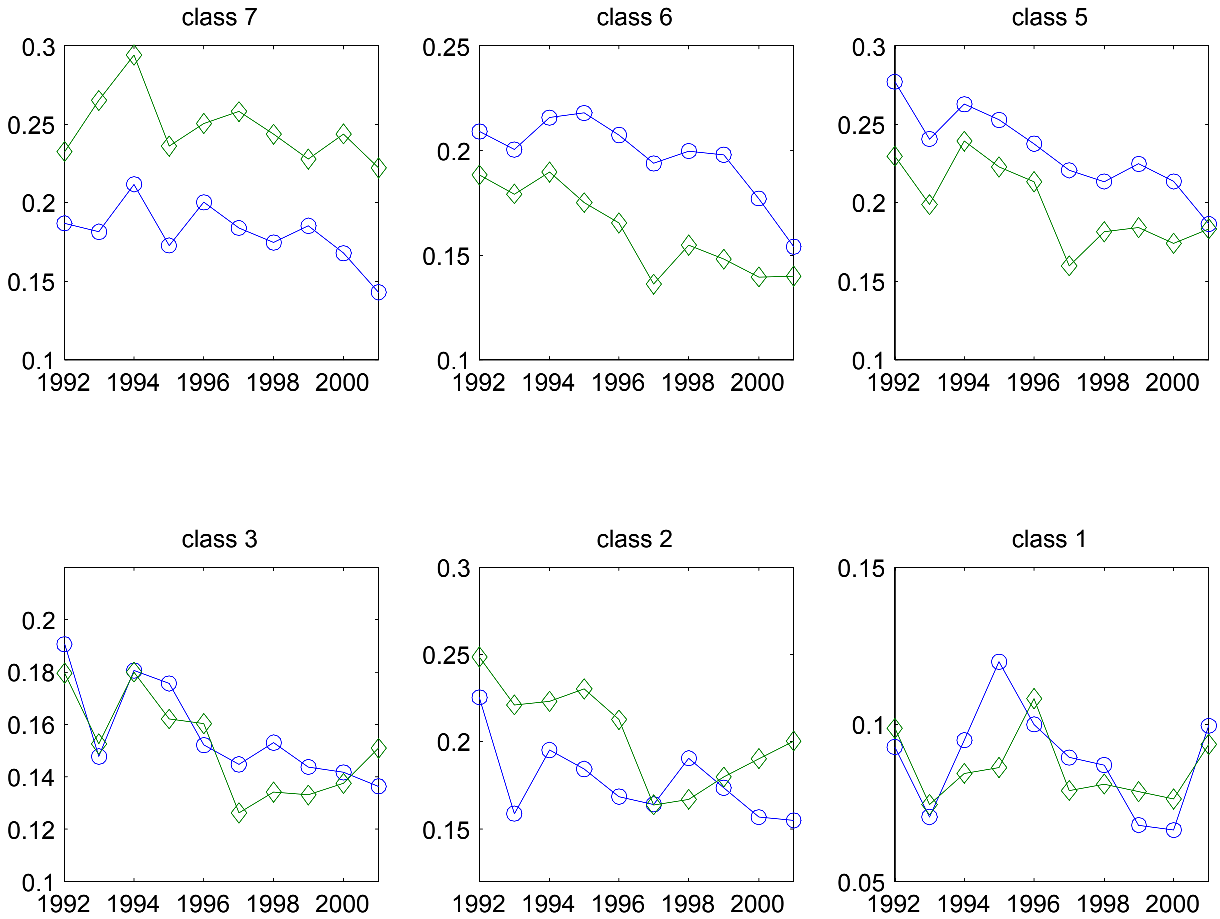

Figure 3: Probabilities of a downgrade of at most 2 buckets for the different rating classes. The circles (resp. the diamonds) correspond to the wholesale (retail trade) sector.

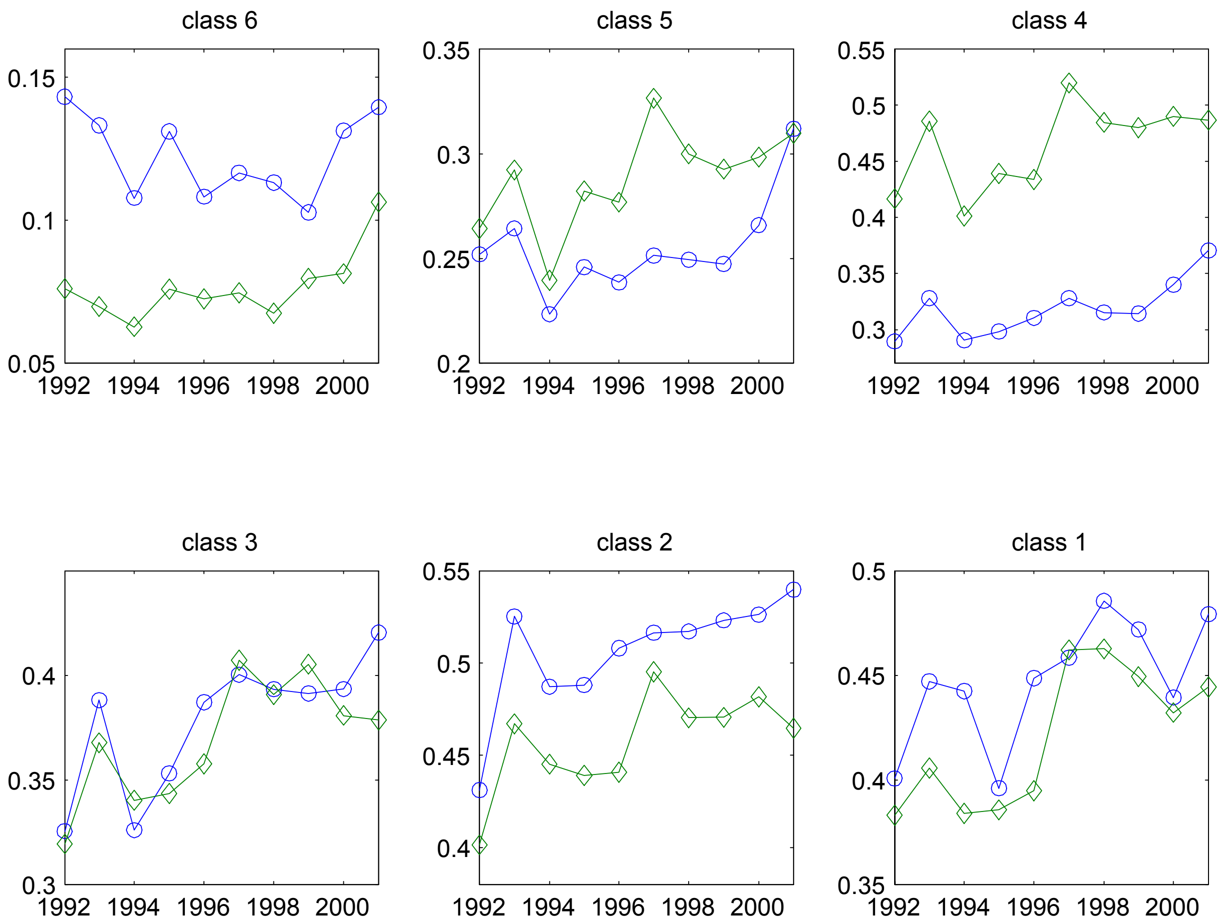

Figure 4: Probabilities of an upgrade of at most 2 buckets for the different rating classes. The circles (resp. the diamonds) correspond to the wholesale (retail trade) sector.

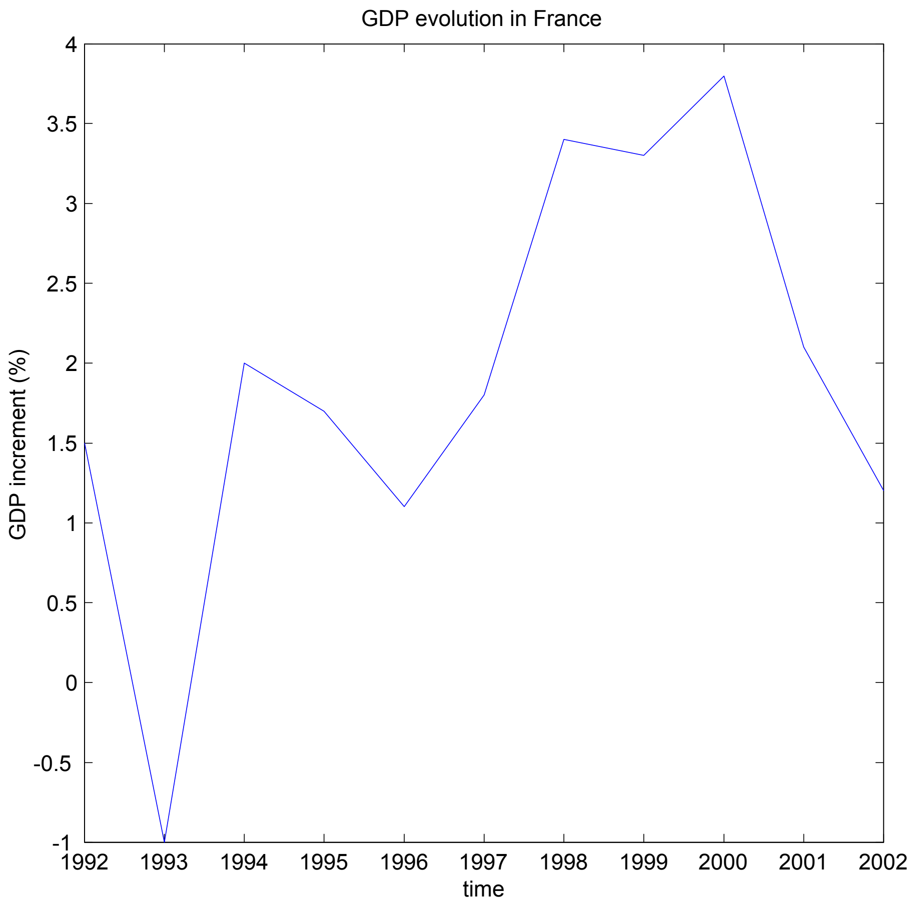

Figure 5: GDP evolution in France in the period 1992-2002.

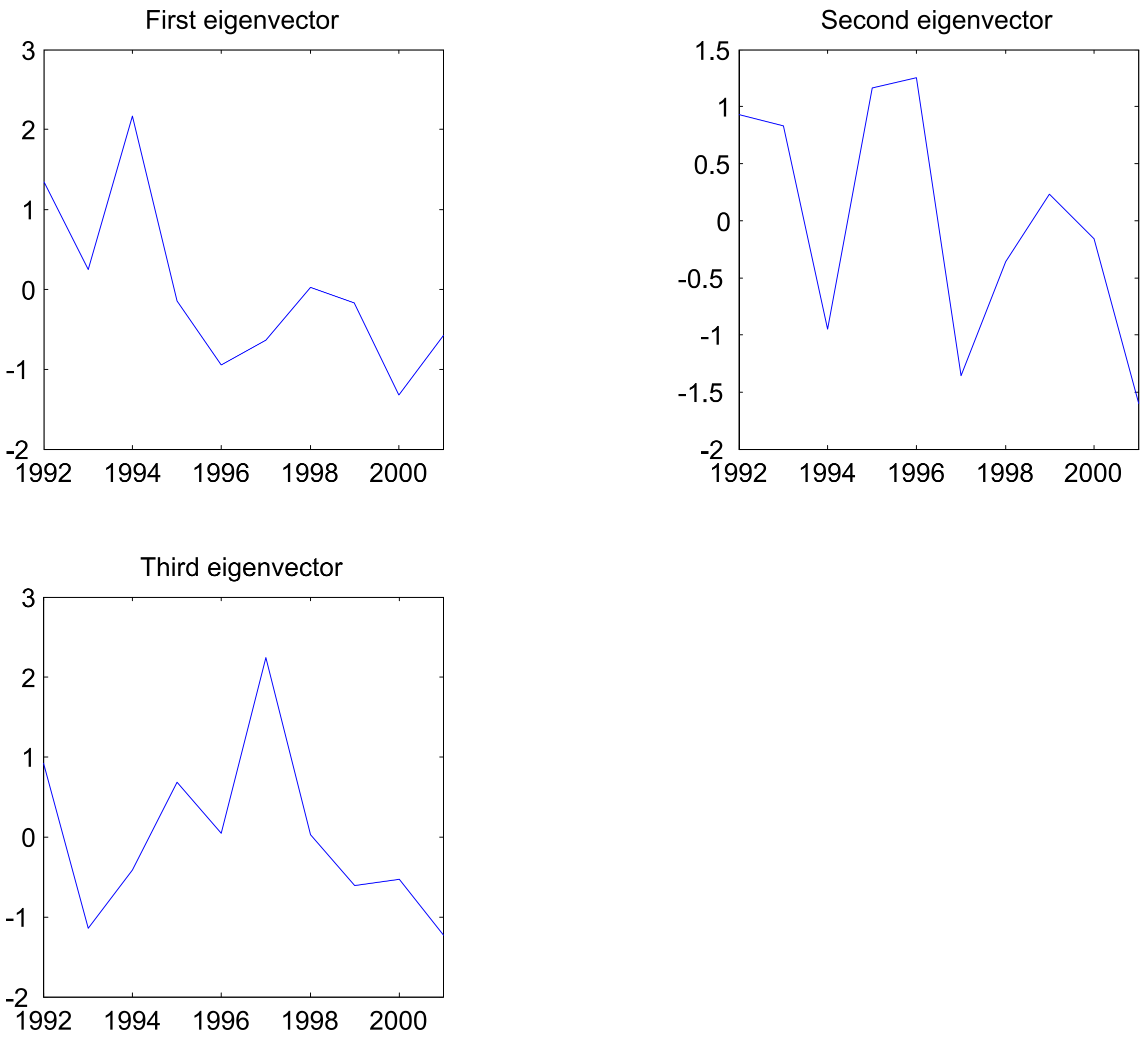

Figure 6: Factors corresponding to the three largest eigenvalues.

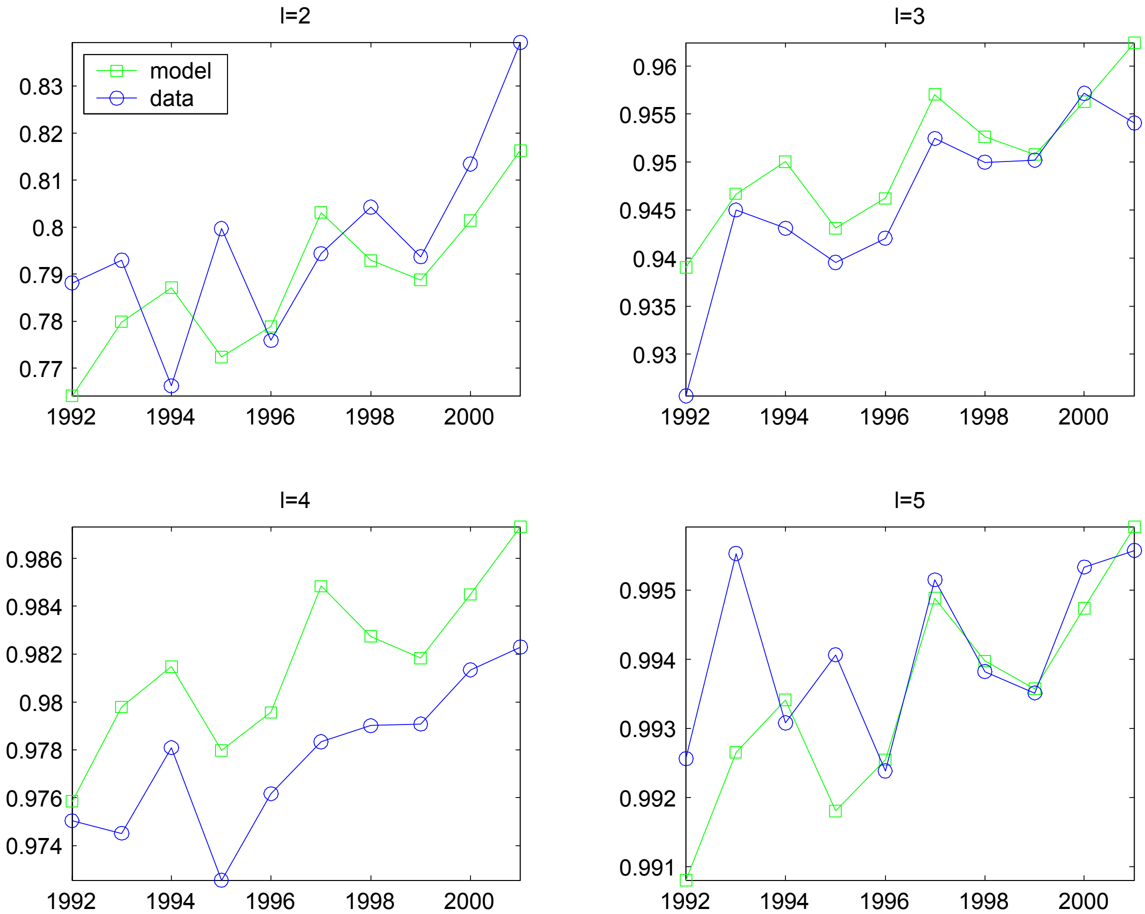

Figure 7: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 1$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.

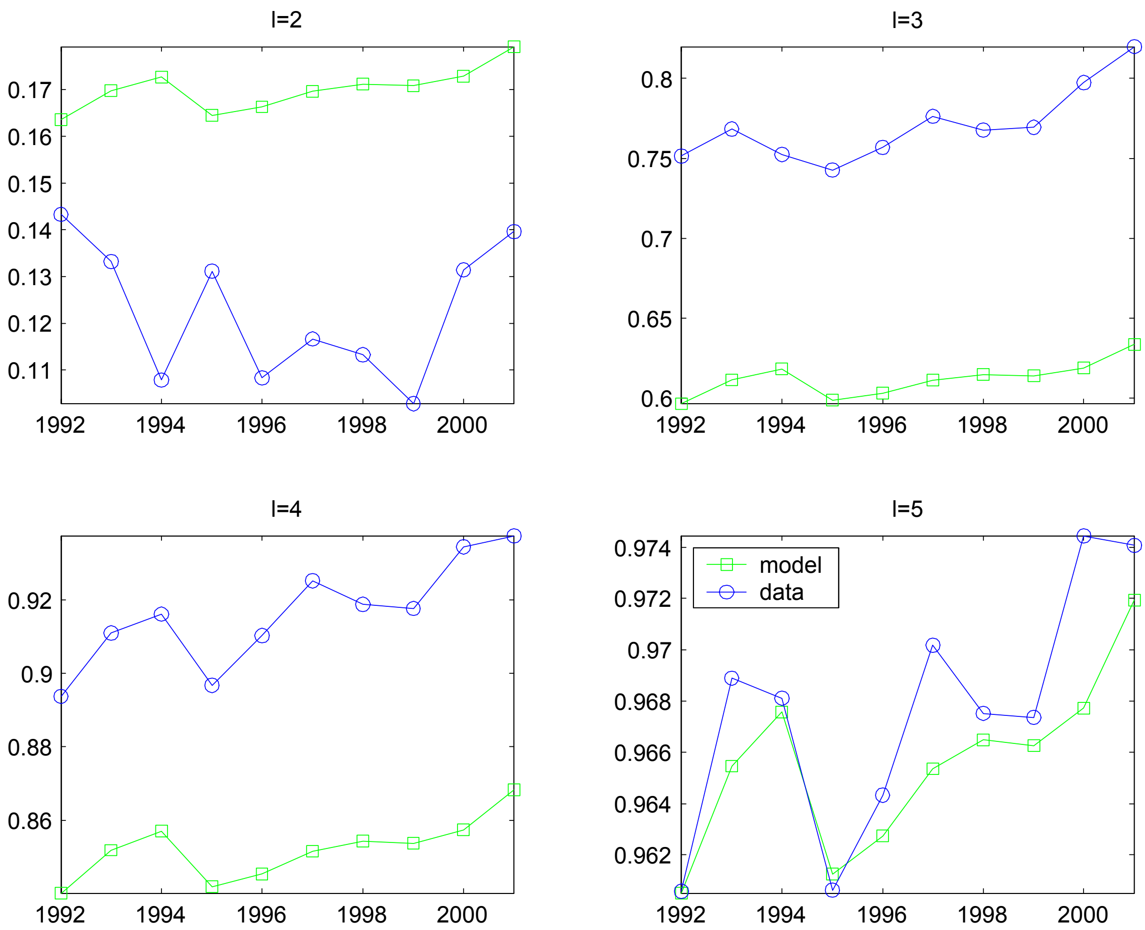

Figure 8: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 2$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.

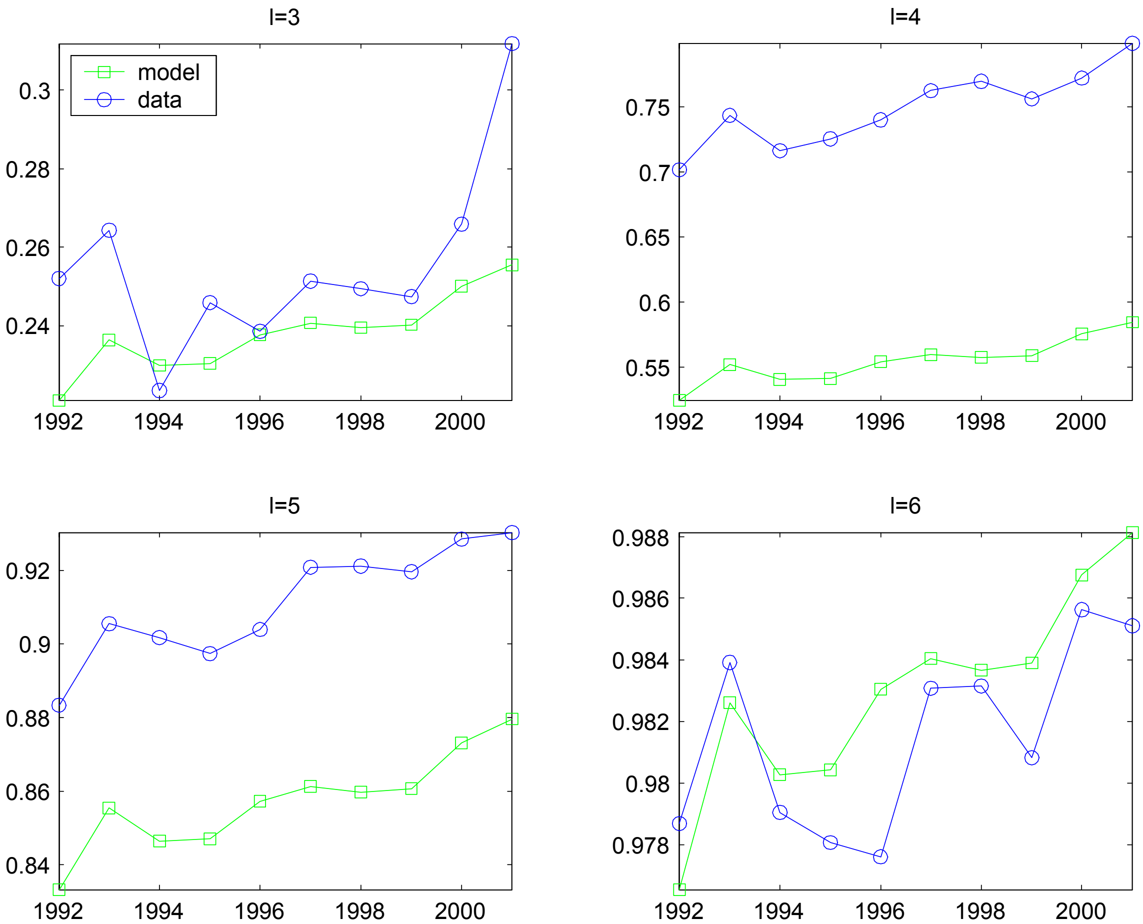

Figure 9: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 3$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.

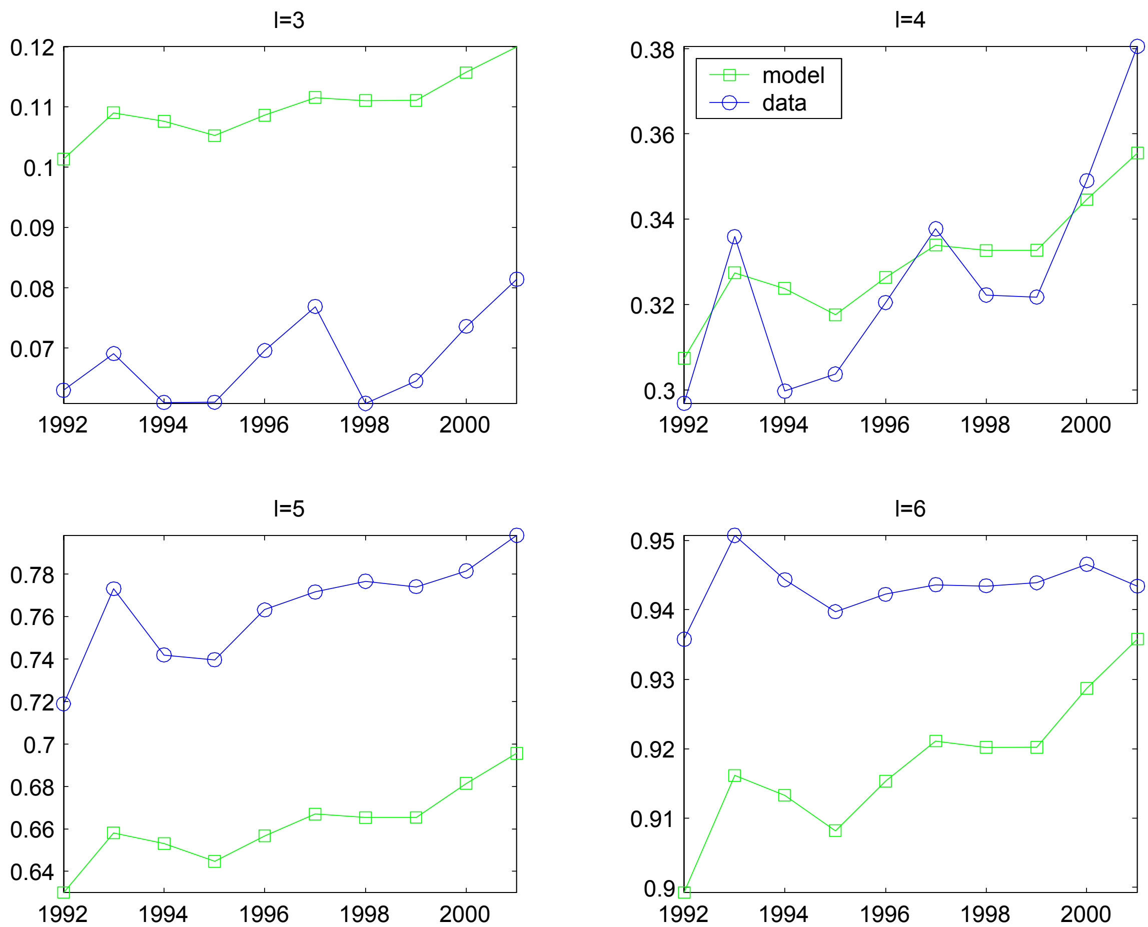

Figure 10: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 4$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.

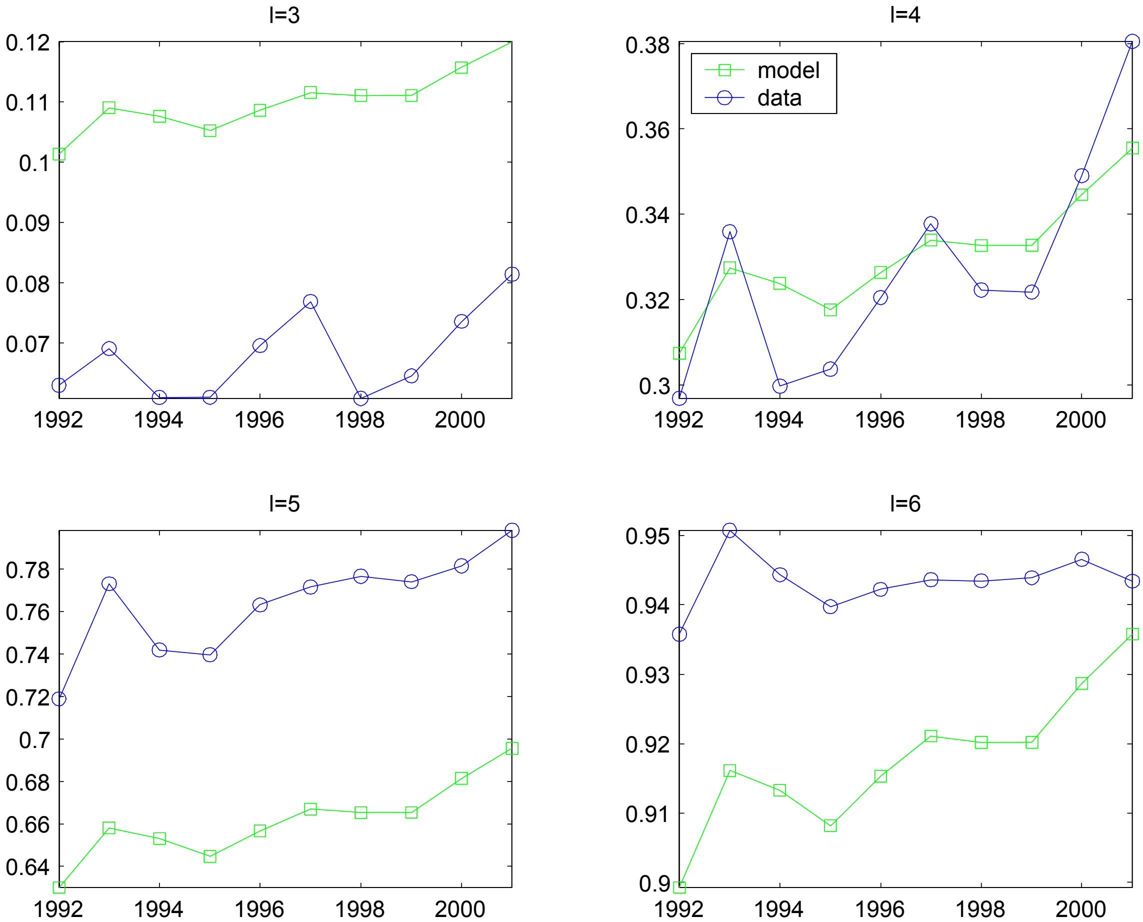

Figure 11: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 5$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.

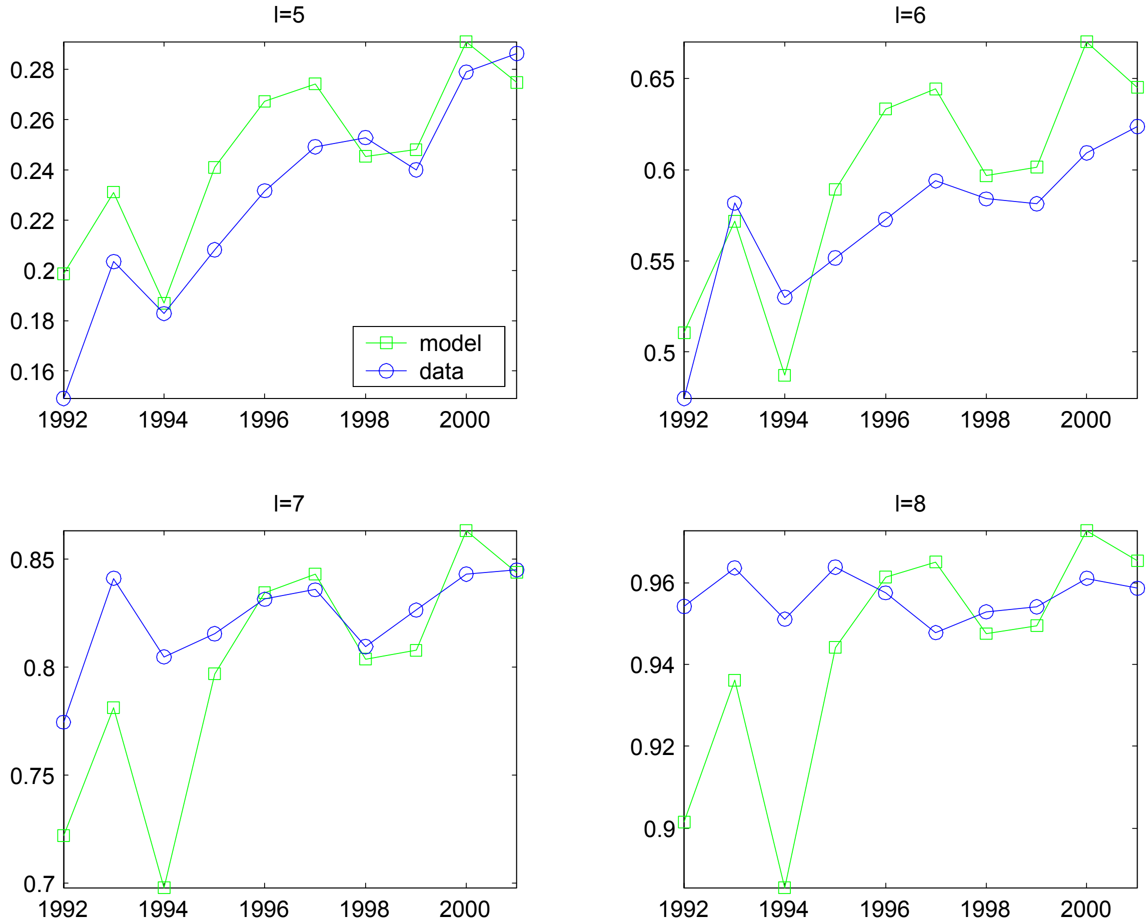

Figure 12: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 6$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.

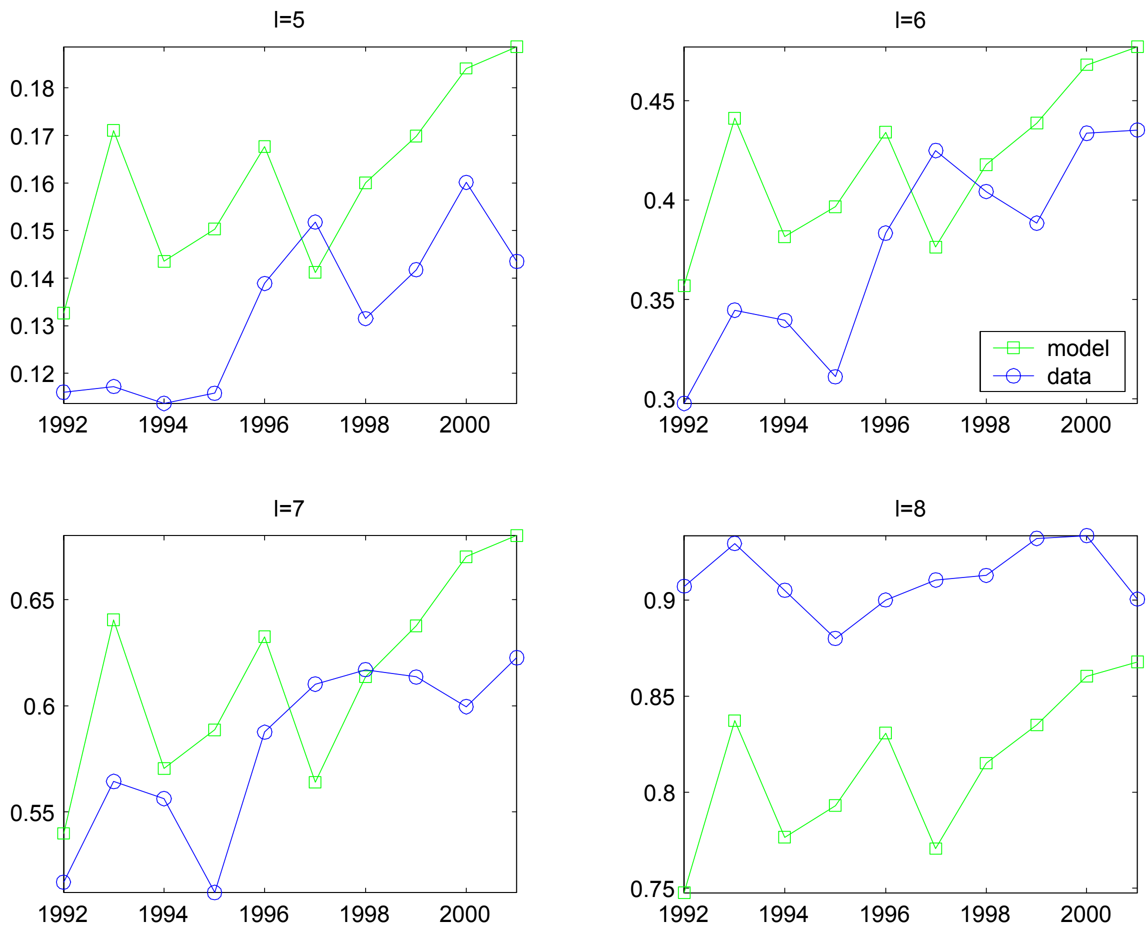

Figure 13: Cumulated probabilities $\pi^*_{kl,t}$ for rating class $k = 7$ and different indices l in the wholesale sector: circles correspond to observed probabilities and squares to fitted probabilities.