Forward Ordinal Probability Models for Point-in-Time Probability of Default Term Structure

Yang, Bill Huajian

September 2017

FORWARD ORDINAL PROBABILITY MODELS FOR POINT-IN-TIME PROBABILITY OF DEFAULT TERM STRUCTURE*

-Methodologies and implementations for IFRS9 ECL estimation and CCAR stress testing

Bill Huajian Yang

Abstract

Common ordinal models, including the ordered logit model and the continuation ratio model, are structured by a common score (i.e., a linear combination of a list of given explanatory variables) plus rank specific intercepts. Sensitivity with respect to the common score is generally not differentiated between rank outcomes. In this paper, we propose an ordinal model based on forward ordinal probabilities for rank outcomes. The forward ordinal probabilities are structured by, in addition to the common score and intercepts, the rank and rating (for a risk-rated portfolio) specific sensitivity. This rank specific sensitivity allows a risk rating to respond to its migrations to default, downgrade, stay, and upgrade accordingly. An approach for parameter estimation is proposed based on maximum likelihood for observing rank outcome frequencies. Applications of the proposed model include modeling rating migration probability for point-in-time probability of default term structure for IFRS9 expected credit loss estimation and CCAR stress testing. Unlike the rating transition model based on Merton model, which allows only one sensitivity parameter for all rank outcomes for a rating, and uses only systematic risk drivers, the proposed forward ordinal model allows sensitivity to be differentiated between outcomes and include entity specific risk drivers (e.g., downgrade history or credit quality changes for an entity in last two quarters can be included). No estimation of the asset correlation is required. As an example, the proposed model, benchmarked with the rating transition model based on Merton model, is used to estimate the rating migration probability and probability of default term structure for a commercial portfolio, where for each rating the sensitivity is differentiated between migrations to default, downgrade, stay, and upgrade. Results show that the proposed model is more robust.

Keywords: ordinal model, forward ordinal probability, common score, rank specific sensitivity, rating migration probability

1. Introduction

Let R denote the outcome for a trial with exactly one of the ordinal outcome values $\{1,2,...,k\}$. The forward ordinal probability, for a rank value i, is the conditional probability that the outcome value is i, given that all outcome ranks are no less than i. While for the backward ordinal probability for a rank value i is the conditional probability that the outcome value is i, given that all outcomes are not larger than i.

Common ordinal models, as reviewed in section 2, include the ordered logit model (i.e., the proportion odd model), and the continuation ratio model. For an ordered logit model, the cumulative probabilities for rank outcomes are modeled by a common score, i.e. a linear combination of a list of explanatory variables, together with rank specific intercept. While for a continuation ratio model, the forward or backward ordinal probabilities for rank outcomes are modeled by a common score with rank specific intercept. Sensitivity with respect to the common score is generally not differentiated between rank outcomes.

It is commonly observed that entities with high risk ratings are more sensitive and vulnerable to adverse shocks, and that entities are more likely to migrate to higher risk grades in the downturn time than to lower risk ratings. Risk sensitivity is generally not uniform between risk ratings, and nor between outcome ranks.

In this paper, we propose an ordinal model based on forward ordinal probability (model (3.2) or (3.4), see section 3). The forward ordinal probabilities are structured by a common score plus rank specific sensitivity and intercept. An algorithm for parameter estimation is proposed based on maximum likelihood approach for observing rank outcome frequencies. The model can be implemented easily by a modeller using, for example SAS PROC NLMIXED ([12]).

Bill Huajian Yang, Enterprise Stress Testing, Royal Bank of Canada, 155 Wellington Street West, Toronto, Canada, M5V 3H6. The views expressed in this article are not necessarily those of Royal Bank of Canada or any of its affiliates. Please direct any comments to bill.yang@rbc.com, phone: 416-313-1245

Applications of the proposed model include: (a) modeling rating migration probability for CCAR stress testing ([2]), and point-in-time probability of default term structure for IFRS9 expected credit loss estimation ([1]); (b) estimation of probability of default for a low default portfolio, and shadow rating modeling.

The modeling of state transition probabilities dates back to the original credit portfolio approaches of CreditMetrics, CreditPortfolioView, and CreditRisk+ ([3], [4]), and contributions by researchers, including the works by Nyström and Skoglund ([8]) and Wei ([11]). Point-in-time rating transition probability model based on Merton model ([5], [6], [7], [10]), structured by a common credit index, is proposed by Miu and Ozdemir ([7]), and is extended by Yang and Du ([13])) to facilitate rating level sensitivity for CCAR stress testing and IFRS9 expected credit loss estimation.

The proposed ordinal model, structured by a common score plus outcome rank specific sensitivity, has several advantages. The outcome rank specific sensitivity allows a risk rating to respond to its migrations to default, downgrade, stay, and upgrade accordingly. Under this model structure, risk for an entity is driven by the common score (as a dynamic) plus sensitivity in responding to a scenario. Unlike the rating transition model ([13]) based on the Merton model framework, which allows only one sensitivity parameter for all outcomes for a rating and uses only systematic risk drivers, this proposed model can include entity specific risk drivers and allows rank specific sensitivity. No estimation for asset correlation is required. Furthermore, the log likelihood based on forward probability of default given by a CDF function is generally concave, greatly increasing optimization efficiency.

Entity specific drivers, such as downgrade history or credit quality changes in the last two quarters, can help improve the prediction and address the issue of the Markov assumption for most migration models, particularly when the portfolio is small and idiosyncratic risk cannot be diversified away.

The paper is organized as follows: In section 2, we review two of the commonly used ordinal regression models: the ordered logit model and the continuation ratio model. In section 3, we propose the forward ordinal model and show the log likelihood function and its concavity. A heuristic hard EM (expectation maximization) algorithm for parameter estimation is proposed in section 4. The model is validated and used in section 5 to model the rating migration probability for a commercial portfolio, where for each rating the sensitivity is differentiated between migrations to default, downgrade, stay, and upgrade. The model is benchmarked with the rating transition model based on Merton framework.

2. A Review of Ordinal Regression Models

In this section, we review two commonly used ordinal models: ordinal regression and continuation ratio models.

Let R denote the outcome for a trial with exactly one of the ordinal outcome values $\{1,2,...,k\}$. Given a scenario consisting of a list explanatory variables $x_1, x_2, ..., x_m$, let $x=(x_1, x_2, ..., x_m)$ denote the corresponding vector. Let $F_i(x)$ and $p_i(x)$ denote, respectively, the cumulative and marginal probabilities defined by

$$F_i(x) = P(R \le i \mid x),$$
$$p_i(x) = P(R = i \mid x)$$

Given x and rank value i, the forward ordinal probability $\tilde{p}_i(x)$ and the backward ordinal probability $\tilde{p}_{bi}(x)$ are defined respectively by the conditional probabilities below

$$\tilde{p}_i(x) = P(R = i \mid x, R \ge i), \quad \tilde{p}_{bi}(x) = P(R = i \mid x, R \le i)$$

Remark. We can always model the backward ordinal probability via the forward ordinal probability model: simply reverse the order of the ordinal outcomes and re-index the resulting forward ordinal probability

$\tilde{p}_i(x)$ by replacing i with $(k+1-i)$. For this reason, we focus our discussion only on forward ordinal probability model. All discussions for forward ordinal model apply naturally to the backward ordinal model by an appropriate reversion for the outcome order and the index of the forward probability.

Proposition 2.1. The following equations hold

$$F_i(x) = p_1(x) + p_2(x) + ... + p_i(x) \quad \text{(2.1A)}$$
$$\tilde{p}_i(x) = p_i(x)/(1 - F_{i-1}(x)) \quad \text{(2.1B)}$$
$$p_i(x) = F_i(x) - F_{i-1}(x) = [(1 - F_{i-1}(x)]\tilde{p}_i(x) \quad \text{(2.1C)}$$
$$[1 - F_i(x)] = [1 - \tilde{p}_1(x)][1 - \tilde{p}_2(x)]...[1 - \tilde{p}_i(x)] \quad \text{(2.1D)}$$

Proof. Equation (2.1A) is immediate. Equation (2.1B) follows from the Bayesian theorem, while equation (2.1C) follows from (2.1A) and (2.1B). By (2.1C), we have

$$\begin{aligned} 1 - F_i(x) &= [1 - F_{i-1}(x)] - p_i(x) \\ &= [(1 - F_{i-1}(x)][1 - \tilde{p}_i(x)] \end{aligned}$$

Thus the last equation (2.1D) follows by induction. □

For the largest rank outcome k, we have the following

$$F_k(x) = 1; \ \tilde{p}_k(x) = 1;$$
$$p_k(x) = 1 - [p_1(x) + p_2(x) + ... + p_{k-1}(x)]$$

Therefore, by Proposition 2.1, an ordinal model can choose to model one of the components below:

(a) the cumulative probabilities $\{F_i(x) \mid i = 1,2,...,k-1\}$
(b) the marginal probabilities $\{p_i(x) \mid i = 1,2,...,k-1\}$
(c) the forward ordinal probabilities $\{\tilde{p}_i(x) \mid i = 1,2,...,k-1\}$.

Marginal probabilities are subject to constraints below

$$p_1(x) + p_2(x) + ... + p_i(x) \le 1,$$
$$p_1(x) + p_2(x) + ... + p_k(x) = 1,$$

Therefore, modeling marginal probabilities individually exposes to additional complexity. In general, one can choose to model either the cumulative probabilities or forward ordinal probabilities, as reviewed and discussed in subsequent sections 2.1 and 2.2.

2.1. Ordinal regression models

An ordinal regression model is generally structured by cumulative probabilities $\{F_i(x) \mid i = 1,2,...,k-1\}$ as

$$F_i(x) = F(b_i + a_1x_1 + a_2x_2 + ... + a_mx_m), \quad b_1 \le b_2 \le ... \le b_{k-1} \quad \text{(2.2)}$$

where F denotes the cumulative distribution for a probability distribution. The coefficients $a_1, a_2, ..., a_m$ in model (2.2) do not depend on index $i \le k-1$.

As cumulative probabilities, $\{F_i(x) \mid i = 1,2,...,k-1\}$ are required to satisfy the following condition

$$F_1(x) \le F_2(x) \le ... \le F_{k-1}(x) \quad \text{(2.3)}$$

This is guaranteed for model (2.2) by the constraint $b_1 \le b_2 \le ... \le b_{k-1}$ in (2.2). Condition (2.3) implies, when modeling the cumulative probabilities, the coefficients $a_1, a_2, ..., a_m$ in (2.2) must be the same for all rank outcomes $\{i = 1, 2, ..., k-1\}$, a limitation for choosing to model the cumulative probabilities.

Recall that, given a sample with n independent trials, where each trial results in exactly one of k rank outcomes, the probability of observing frequencies $\{n_i\}$, with frequency n_i for the i^{th} outcome, is

$$\frac{n!}{n_1! n_2! ... n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}, \quad n = n_1 + n_2 + ... + n_k$$

where $p_i = p_i(x)$ is the marginal probability for rank outcome i, which can be derived from the cumulative probabilities given in (2.2). Therefore, the parameters for model (2.2) can be estimated by using the maximum likelihood approaches, given a sample for the observed rank outcome frequencies.

The proportion odd (or ordered logistic regression) model, a commonly used ordinal model, is given by

$$\begin{aligned} &\log[P(R \le i \mid x) / P(R > i \mid x)] = b_i + a_1 x_1 + a_2 x_2 + ... + a_m x_m \\ &\Rightarrow F_i(x) = P(R \le i \mid x) \\ &\qquad = 1/[1 + \exp(-b_i - a_1 x_1 - a_2 x_2 - ... - a_m x_m)] \\ &\qquad = F(b_i + a_1 x_1 + a_2 x_2 + ... + a_m x_m)] \end{aligned}$$

where $F(x) = 1/(1+\exp(-x))$ is the standard logistic cumulative probability distribution. Thus the proportion odd model is a special case of the ordinal regression model (2.2) with the link function given by the inverse of the standard logistic cumulative distribution, i.e., the logit function.

Ordinal regression models are implemented by SAS, with options for different link functions, including the inverse of standard logistic and the inverse of standard normal cumulative distributions (i.e., the logit and probit functions).

2.2. Forward/backward continuation ratio model

Recall that the logit function is defined as $\text{logit}\,(p) = \log[\,p/(1\text{-}p)]$ for $0 < p < 1$. The forward and backward logistic continuation ratio models are structured, respectively, by equations (2.4A) and (2.4B) below, given scenario x and rank outcome value i:

$$\text{logit}\,[P(R = i \mid x) / P(R \ge i \mid x)] = b_i + a_1 x_1 + a_2 x_2 + ... + a_m x_m \qquad (2.4A)$$

$$\text{logit}\,[P(R = i \mid x) / P(R \le i \mid x)] = b_i + a_1 x_1 + a_2 x_2 + ... + a_m x_m \qquad (2.4B)$$

The coefficients $a_1, a_2, ..., a_m$ do not depend on index $i \le k-1$. Let $\tilde{p}_i(x)$ denote the forward ordinal probability $P(R = i \mid x, R \ge i)$ or the backward ordinal probability $P(R = i \mid x, R \le i)$. Then we can reformulate (2.4A) and (2.4B) as

$$\begin{aligned} \tilde{p}_i(x) &= 1/(1 + \exp(b_i + a_1 x_1 + a_2 x_2 + ... + a_m x_m)) \\ &= \Phi(b_i + a_1 x_1 + a_2 x_2 + ... + a_m x_m) \end{aligned} \qquad (2.5)$$

where Φ denotes standard logistic cumulative distribution. This means the logistic forward continuation ratio model is structured by the forward ordinal probabilities for rank outcomes, with the inverse of the standard logistic cumulative distribution, i.e., the logit function, as the link function. The probit continuation ratio model is structured similarly using the inverse of the standard normal cumulative distribution, i.e., the probit function, as the link function.

3. The Proposed Forward Ordinal Model

With ordinal regression model (2.2) and continuation ratio models (2.4A)-(2.4B), sensitivities for all rank outcomes are all the same, though the intercept can be different between rank outcomes. In this section, we propose an ordinal model based on forward ordinal probabilities. This forward ordinal model allows sensitivity to be differentiated between rank outcomes.

3.1. The mathematical setup

We assume, given that the rank outcome will be no less than i, i.e., $R \geq i$, there is a latent variable y_i given by

$$y_i = -b_i - r_i(a_1x_1 + a_2x_2 + ... + a_mx_m) + \varepsilon_i \tag{3.1}$$

such that the outcome $R > i$ when $y_i > 0$; and $R = i$ if $y_i \leq 0$; where ε_i is a random variable with zero mean, independent of $x = (x_1, x_2, ..., x_m)$. The coefficients $\{a_1, a_2, ..., a_m\}$ do not depend on index $i \leq k-1$.

By an appropriate scaling to both sides of (3.1), we can assume the standard deviation of ε_i is 1. We assume that ε_i is standard normal. Let Φ denote the cumulative distribution for ε_i. Then the forward ordinal probability $\tilde{p}_i(x)$ by (3.1) is

$$\tilde{p}_i(x) = \Phi(b_i + r_i(a_1x_1 + a_2x_2 + ... + a_mx_m)) \tag{3.2}$$

Let $c(x) = (a_1x_1 + a_2x_2 + ... + a_mx_m)$. We call $c(x)$ a common score and r_i the sensitivity for the rank value $i \leq k-1$ with respect to the common score $c(x)$. For IFRS9 expected loss estimation and CCAR stress testing, $c(x)$ can include both systematic and entity specific risk drivers.

Note that, with model (3.2), an increase (resp. decrease) for the norm of the parameter vector $(a_1, a_2, ..., a_m)$ during parameter estimation can propagate to the sensitivity parameter vector $(r_1, r_2, ..., r_{k-1})$ by a scale down (resp. up). To prevent unnecessary disturbance of parameter estimation and ensure estimation convergence, the following constraints can be imposed

$$a_1{}^2 + a_2{}^2 + ... + a_m{}^2 = 1 \tag{3.3A}$$

In practice, the sign of a coefficient a_i is pre-determined. For example, default risk increases as unemployment rate increases. We thus require the coefficient for unemployment rate in the model to be positive. In this case, we can assume that all $\{a_i\}$ are nonnegative by an appropriate sign scaling to the corresponding variable. Then a linear constraint as below can be imposed

$$a_1 + a_2 + ... + a_m = 1 \tag{3.3B}$$

Let $c(x) = (a_1x_1 + a_2x_2 + ... + a_mx_m)$. In the case when variables $x_1, x_2, ..., x_m$ are common to all entities (e.g., the macroeconomic variables), we have the model (3.4) below, assuming the normality for $c(x)$ with mean u and standard deviation v:

$$\tilde{p}_i(x) = \Phi(c_i\sqrt{1 + (r_iv)^2} + r_i(a_1x_1 + a_2x_2 + ... + a_mx_m - u)) \tag{3.4}$$

where c_i is the threshold value estimated directly by taking the inverse Φ^{-1} to the long-run average for forward ordinal probability, which can be estimated directly from the sample. Model (3.4) is derived from (3.2) by a well-known lemma ([9]) for the expectation with respect to s

$$E_s[\Phi(a+bs)] = \Phi(a/\sqrt{1+b^2}),\ s \sim N(0,1)$$

With model (3.4), estimation is required only for parameters $\{a_1, a_2, ..., a_m\}$ and $\{r_i\}$, not the intercepts $\{b_i\}$.

3.2. The log-likelihood function given the observed rank frequencies

In this section, we show the log-likelihood and its concavity for observing rank outcome frequencies by using the forward ordinal probabilities $\{\tilde{p}_i(x) | i = 1, 2, ..., k\}$.

Given a scenario $x = (x_1, x_2, ..., x_m)$, let n_i denote the corresponding observed frequency for the i^{th} rank value. Let

$$n = n_1 + n_2 + ... + n_k \tag{3.5A}$$

Define s_i by

$$s_i = n - (n_1 + n_2 + ... + n_{i-1}) = n_k + n_{k-1} + ... + n_i \tag{3.5B}$$

We focus on the conditional probability space given that the rank value of the outcome R is no less than i. The log-likelihood for observing frequency n_i for the i^{th} rank value and frequency $s_i - n_i$ for rank values larger than i, given $x = (x_1, x_2, ..., x_m)$, is

$$L_i(x) = (s_i - n_i)\log[1 - \tilde{p}_i(x)] + n_i \log[\tilde{p}_i(x)] \tag{3.6}$$

up to a summand given by the logarithm of a binomial coefficient, which is independent of model parameters of model (3.2) and (3.4), assuming the occurrence of the i^{th} rank value is a binary event.

Let $L(x, i, i+h)$ denote the log likelihood over this probability space for observing multiple frequencies $\{n_i, n_{i+1}, ..., n_{i+h}\}$ for rank values $\{i, i+1, ..., i+h\}$ and the frequency

$$s_{i+h+1} = n_k + n_{k-1} + ... + n_{i+h+1}$$

for rank values larger than $i+h$. We have the following proposition (see Appendix for a proof).

Proposition 3.1. Equations (3.7A) and (3.7B) hold up to a summand given by the logarithms of some binomial coefficients (independent of the parameters in model (3.2) and (3.4)):

$$L(x, i, i+h) = L_i(x) + L_{i+1}(x) + ... + L_{i+h}(x) \tag{3.7A}$$

$$L(x, 1, k) = L_1(x) + L_2(x) + ... + L_k(x) \tag{3.7B}$$

□

A function is log concave if its logarithm is concave. If a function is concave, a local maximum is a global maximum, and the function is unimodal. This property is important for maximum likelihood estimate search. A proof for the proposition below can be found in Appendix.

Proposition 3.2. The log likelihood function (3.7A) and (3.7B), with Φ being the standard normal cumulative probability distribution, is concave in the following two cases:

(a) As a function of the r-parameters $\{r_i\}$, or of the b-parameters $\{b_j\}$, and the a-parameters $\{a_1, a_2, ..., a_m\}$ when $\tilde{p}_i(x)$ is given by (3.2).

(b) As a function of the a-parameters $\{a_1, a_2, ..., a_m\}$, or as a function of the r-parameters $\{r_i\}$ when $\tilde{p}_i(x)$ is given by (3.4).

□

4. Parameter Estimation by Maximum Likelihood Approaches

In this section, we propose an algorithm for parameter estimation for models (3.2) and (3.4) by maximizing the log likelihood for observing rank outcome frequencies. This generic algorithm works for one forward ordinal. For modeling rating migration for a risk rated portfolio, multiple forward ordinal models are required, with one for each of non-default risk rating (See section 5 for model formulation and the adapted algorithm for parameter fitting).

A. Estimation of parameters for model (3.2)

The algorithm proposed is essentially a heuristic hard EM (expectation maximization) algorithm.

Parameter initialization: Initially, $\{r_1, r_2, ..., r_{k-1}\}$ are set to 1. Estimate the parameters $\{a_1, a_2, ..., a_m\}$ and $\{b_1, b_2, ..., b_{k-1}\}$, without the constraint (3.3A) and (3.3B), by maximizing the log-likelihood of (3.7B). Recall that (3.7B) is concave by Proposition 3.2 (a), therefore global maximum estimates are granted. Rescale the a-parameter estimates by a scalar $\rho > 0$ to make $(a_1, a_2, ..., a_m)$ a unit vector, and then set $\{r_1, r_2, ..., r_{k-1}\}$ each to $1/\rho$. This completes the initialization for all parameters.

Step 1. Assume that the sensitivities $\{r_1, r_2, ..., r_{k-1}\}$ and $\{b_1, b_2, ..., b_{k-1}\}$ are given. Estimate the parameters $\{a_1, a_2, ..., a_m\}$ by maximizing the log-likelihood of (3.7B). Global maximum estimates are granted by Proposition 3.2 (a).

Step 2. Assume that the parameters $\{a_1, a_2, ..., a_m\}$ and $\{b_1, b_2, ..., b_{k-1}\}$ are given. Estimate the sensitivities $\{r_1, r_2, ..., r_{k-1}\}$ by maximizing the log-likelihood of (3.7B). Recall that, by Proposition 3.2 (a), global maximum estimates are granted.

Step 3. Assume that the parameters $\{a_1, a_2, ..., a_m\}$ and $\{r_1, r_2, ..., r_{k-1}\}$ are given. Estimate the sensitivities $\{b_1, b_2, ..., b_{k-1}\}$ by maximizing the log-likelihood of (3.7B). Also by Proposition 3.2 (a), global maximum estimates are granted.

Step 4. Iterate the above three steps until a convergence is reached. Steps 1-3 are repeated until convergence is reached, i.e., the maximum deviation for all parameter estimates for $\{b_1, b_2, ..., b_{k-1}\}$, $\{a_1, a_2, ..., a_m\}$, and $\{r_1, r_2, ..., r_{k-1}\}$ in consecutive two iterations, is less than 10^{-4}.

We implement the above three-step optimization process by using the SAS procedure PROC NLMIXED.

B. Estimation of parameters for model (3.4)

For model (3.4), follows steps 1-4 above to fit for the coefficients $\{a_1, a_2, ..., a_m\}$ for common score $c(x) = (a_1 x_1 + a_2 x_2 + ... + a_m x_m)$. When this common score is known, we estimate $\{r_1, r_2, ..., r_{k-1}\}$ by maximizing (3.7B) with $\tilde{p}_i(x)$ being given by (3.4). Global maximum estimates are granted Proposition 3.2 (b).

5. An Empirical Example: Rating Migration Probability and PD Term Structure for a Commercial Portfolio

In this section, we apply the proposed ordinal model to estimate the rating transition probability for a risk rated commercial portfolio. Point-in-time PD term structure, for IFRS9 ECL estimation and CCAR stress testing, is derived.

The sample contains quarterly rating migration frequencies between 2006Q3 and 2016Q4 for a commercial portfolio, created synthetically by scrambling the default rate by an appropriate scaling. There are 21 risk ratings, with R_{21} as the default rating, and R_1 the best quality rating.

Because we are more concerned with the default outcome and default risk, we model rating migration probability by backward ordinal model, starting with the most important rating level default risk. As noted in section 2, a backward ordinal model can be viewed as a forward ordinal model after an appropriate reversion of the outcome order and the index of the resulting forward ordinal probability.

The backward ordinal model is benchmarked with the rating transition model based on Merton model proposed by Yang and Du ([13]). Additional benchmark comments for SAS ordinal regression using SAS PROC LOGISTIC are given at the end of the section.

5.1. The backward ordinal and benchmark models for IFRS9 expected credit loss estimation and CCAR stress testing

A. Formulation of the models

(a) Backward ordinal model for rating migration probability

Given a non-default initial risk rating R_i at the beginning of the quarter, there are 21 possible ordinal outcomes at the end of the quarter: an entity can migrate to default rating or any of the other 20 ratings. Given a scenario $x = (x_1, x_2, ..., x_m)$, let $\tilde{p}_{ij}(x)$ denote the backward ordinal probability that the rating R_i migrates to rating R_j given that it will migrate only to a rating with rank no larger than j. Bearing in mind that a backward ordinal model can be viewed as a forward ordinal model by an outcome order and probability index reversion, we can model $\tilde{p}_{ij}(x)$ by models (3.2) and (3.4) as (5.1A) and (5.1B) respectively:

$$\tilde{p}_{ij}(x) = \Phi(b_{ij} + r_{ij}(a_1x_1 + a_2x_2 + ... + a_mx_m)) \tag{5.1A}$$

$$\tilde{p}_{ij}(x) = \Phi(c_{ij}\sqrt{1+(r_{ij}v)^2} + r_{ij}(a_1x_1 + a_2x_2 + ... + a_mx_m - u)) \tag{5.1B}$$

We assume that, for each initial rating R_i, the sensitivity parameter r_{ij} are the same for rank outcome values j when: (a) $i < j < 21$ (downgrade); (b) $1 \le j < i$ (upgrade). Denote the downgrade sensitivity r_{id} and the upgrade sensitivity by r_{iu}. Let r_{idf} and r_{is} be the sensitivities respectively for outcome cases (c) $j = 21$ (default); and (d) $j = i$ (stay). Then (5.1A) and (5.1B) reduces to (5.2A) and (5.2B) below:

$$\tilde{p}_{ij}(x) = \Phi(b_{ij} + r_i(a_1x_1 + a_2x_2 + ... + a_mx_m)) \tag{5.2A}$$

$$\tilde{p}_{ij}(x) = \Phi(c_{ij}\sqrt{1+(r_iv)^2} + r_i(a_1x_1 + a_2x_2 + ... + a_mx_m - u)) \tag{5.2B}$$

where $r_i = r_{idf}, r_{id}, r_{is}, r_{iu}$ respectively for default, downgrade, stay, and upgrade. The marginal probability is given by

$$p_{ij}(x) = (1 - F_{ij}(x))\tilde{p}_{ij}(x)$$

where $F_{ij}(x) = p_{i21}(x) + p_{i20}(x) + ... + p_{i\,22-j}(x)$ is the cumulative probability. Constraint (3.3A) or (3.3B) is imposed for the proposed backward ordinal model (5.2A) and (5.2B).

(b) Rating transition model under the Merton model framework

Point-in-time rating transition probability model based on Merton framework is proposed by Miu and Ozdemir ([7]), and is extended by Yang and Du ([13])) to facilitate rating level sensitivity for CCAR stress testing and IFRS9 expected credit loss estimation.

Let $t_{ij}(x)$ denote the transition probability from an initial rating R_i at the beginning of the quarter to rating R_j at the end of the quarter, given a macroeconomic scenario $x = (x_1, x_2, ..., x_m)$. Let Φ denote the standard normal cumulative distribution. Under the Merton model framework ([5], [6], [7], [10]), it can be shown ([13]) that

$$\begin{aligned} t_{ij}(x) &= \Phi(\bar{q}_{i(k-j+1)} + \tilde{r}_i ci(x)) - \Phi(\bar{q}_{i(k-j)} + \tilde{r}_i ci(x)) \\ &= \Phi[\bar{q}_{i(k-j+1)} + \tilde{r}_i(\tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + ... + \tilde{a}_m\tilde{x}_m)] \\ &\quad - \Phi[\bar{q}_{i(k-j)} + \tilde{r}_i(\tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + ... + \tilde{a}_m\tilde{x}_m)] \end{aligned} \qquad (5.3)$$

where $\bar{q}_{ih} = q_{ih}\sqrt{1+\tilde{r}_i^2}$, the quantities $\{q_{ij}\}$ are the threshold values given by $q_{ij} = \Phi^{-1}(\bar{p}_{ij})$, where $\bar{p}_{ij}$ is the through-the-cycle transition probability from rating R_i to rating R_j, which can be estimated directly from the historical sample. The sensitivity parameter $\tilde{r}_i$ is the same for all rank outcomes for a given rating R_i.

The index $ci(x) = \tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + ... + \tilde{a}_m\tilde{x}_m$ is derived by a normalization from a linear combination $a_1\tilde{x}_1 + a_2\tilde{x}_2 + ... + a_m\tilde{x}_m$, with which the model $\{p_i(x)\}$ best predicts the portfolio default risk, in the sense of maximum likelihood for observing default frequencies, where

$$p_i(x) = \Phi[c_i + \tilde{r}_i(\tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + ... + \tilde{a}_m\tilde{x}_m)] \qquad (5.4)$$

is a model predicting the probability of default for rating R_i, no constraint is imposed for intercept c_i. The quantity $\tilde{r}_i$ is driven by

$$\tilde{r}_i = r_i\lambda / \sqrt{1 + r_i^2(1-\lambda^2)}\ ,\ 0 \le \lambda \le 1 \qquad (5.5)$$

where $r_i = \sqrt{\rho_i} / \sqrt{1-\rho_i}$ and ρ_i is the asset correlation in the Merton model for rating R_i ([13]).

Remark. We can choose to fit for $\{a_1, a_2, ..., a_m\}$ without constraint (5.5). Unconstrained result is always better than the constrained one in the sense of higher likelihood value.

B. Fitting for parameters

We focus on macroeconomic scenarios and consider parameter fitting only for models (5.2B) and (5.3).

For models (5.2B) and (5.3), parameter fitting follows the two steps below:

1. Fit for the macroeconomic variable coefficients $\{a_1, a_2, ..., a_m\}$ by maximum likelihood for observing rating level default frequencies, with default probability $p_i(x)$ for rating R_i being given by (5.4) without constraint (5.5). This can be done similarly as steps 1-4 in section 4.
2. When credit index $ci(x) = \tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + ... + \tilde{a}_m\tilde{x}_m$ is determined, we are required to fit only for the risk sensitivity parameters $\{r_i\}$ for model (5.3), and $\{r_{idf}, r_{id}, r_{is}, r_{iu}\}$ for model (5.2B), for ratings $\{R_i\}$. For model (5.3), we can choose to fit for $\{r_i\}$ either separately for each rating R_i, or in a combined way for all ratings $\{R_i\}$, by using the appropriate likelihood function (3.7B) for all rating

migration frequencies, or (3.7A) for downgrade or default frequencies only. The corresponding log likelihood function is concave by Proposition (3.2) (b). For model (5.2B), we fit for each of the four groups $\{r_{idf}\},\{r_{id}\},\{r_{is}\},\{r_{iu}\}$ separately, using the appropriate likelihood function (3.7A) for the corresponding migration frequency.

In general, monotonicity for sensitivity between ratings is imposed: Specifically, we require that $\{r_i\}$, $\{r_{idf})$ and $\{r_{id}\}$ be non-decreasing and that $\{r_{is}\}$ and $\{r_{iu}\}$ be non-increasing for a higher risk rating.

5.2. Validation results

We use the following labels for the backward ordinal and the benchmark models:

1. BORD – the backward ordinal model (5.2B)
2. RTGM – the rating migration model based on Merton model framework (5.3)

All three models use the same variables as listed below, provided by the US Federal Reserve:

1. 3-month treasury bill interest rate
2. Unemployment rate

The macro coefficients for credit index $ci(x) = a_1 x_1 + a_2 x_2 + ... + a_m x_m$ are fitted as described in section 5.1 in the same way for both the backward ordinal model and rating migration model based on Merton model, so both models have the same macro coefficient estimates. The table below records the estimates for these two coefficients, with the variable p-values p_1 and p_2.

Table 1. Macro coefficients

Model	v1	v2	p1	p2
RTGM/BORD	0.3975	0.6025	0.0247	0001

For the backward ordinal model the sensitivity parameter estimates are reported as in the table below for 20 non-default ratings for default (DF), downgrade (DG), and stay (Stay), with monotonicity constraint being imposed. The sensitivity estimates for upgrade are all close to zero (reflecting that fact that the upgrade probability is slim in the stress period), and are not printed in the table. The migration model based on Merton model estimates the sensitivity parameters by maximum likelihood for observing only the default frequency, thus it has the same sensitivity parameter estimates as the backward ordinal model for default (the first row of the table).

Table 2. Sensitivity parameter estimates

Migration	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DF	0.001	0.001	0.002	0.003	0.004	0.005	0.017	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.060	0.060	0.769	0.769
DG	0.128	0.129	0.130	0.131	0.132	0.133	0.134	0.135	0.136	0.151	0.228	0.229	0.230	0.231	0.248	0.249	0.362	0.504	1.139	.
Stay	1.992	1.992	0.258	0.258	0.258	0.144	0.133	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.050	0.050	0.050	0.050	0.050

The table below show the back-test performance for two models based on R-Squared for prediction of portfolio cumulative default rates for 1, 4, 6, 12, and 16 quarters for the derived point-in-time PD term structure.

Table 3. RSQ for portfolio cumulative default rate

Model	1	4	8	12	16
BORD	0.420	0.575	0.570	0.792	0.777
RTGM	0.420	0.558	0.518	0.726	0.660

The results show model performance improves for the backward ordinal model when sensitivity parameter is differentiated between migrations to default, downgrade, stay, and upgrade for the backward ordinal model. This improvement is a trade-off to adding more sensitivity parameters.

We end this section by adding additional benchmark comment based on SAS ordinal regression using SAS PROC LOGISTIC, with both logit and probit as the link functions, via the "class" and "by" options.

When "by" statement is used for initial ratings, SAS fits for each initial rating R_i an ordinal regression model of the form

$$F_{ij}(x) = \Phi(b_{ij} + a_{i1}x_1 + a_{i2}x_2 + ... + a_{im}x_m)$$

for the cumulative probability for rank outcome j less than 21. This model has redundant coefficients (depending on rating index i), causing an over-fit issue for such a short time series sample. More importantly, it is not structured by a common score, and the sensitivity. We do not recommend this model.

When the "class" statement is used, the initial risk rating is treated as a class variable in the model, and SAS fits for each initial risk rating R_i an ordinal model of the form as

$$F_{ij}(x) = \Phi(b_{ij} + a_1x_1 + a_2x_2 + ... + a_mx_m)$$

for the cumulative probability for the rank outcome j less than 21. The intercept vectors for initial risk rating R_i and R_1 satisfy the following equation:

$$(b_{i1}, b_{i2}, ..., b_{i20}) = (d_i + b_{11}, d_i + b_{12}, ..., d_i + b_{120}) \tag{5.4}$$

with constant d_i corresponding to the i^{th} level of the class variable. That is, the intercept vector for R_i is a translation of the intercept vector for R_1. As expected, this model fails to predict the default risk and other migration risk. It over-estimates PD for the high risk ratings $R_{19,} R_{20}$, and under-estimates significantly the PD for other ratings. We do not recommend this model.

.

Conclusions. Ordinal regression models are widely used for modeling rating migration. Results are generally not very optimistic, partly due to the lack of flexibility with respect to the sensitivity (between rank outcomes and between risk ratings). In this paper, we propose an ordinal model based on forward ordinal probabilities. Under this model, forward ordinal probabilities are structured by a common score plus rank and rating specific sensitivity. This rank specific sensitivity allows a risk rating to respond to its own migration patterns to default, downgrade, stay, and upgrade accordingly. Empirical results show, the model is more robust than the rating transition model based on the Merton model framework. Unlike the rating transition model based on Merton model, which allows only one sensitivity parameter for all rank outcomes for a rating, and uses only systematic risk drivers, the proposed ordinal model differentiate sensitivity between outcomes and include entity specific risk drivers. No estimation for asset correlation is required. The model can be implemented by using for example, the SAS PROC NLMIXED procedure. This forward ordinal model will provide a new and useful tool for practitioners for point-in-time PD term structure modeling for IFRS9 expected credit loss estimation, and multi-period scenario loss projection for CCAR stress testing.

Acknowledgements: The author thanks both referees for many insights and valuable suggestions. Special thanks go to Carlos Lopez for his consistent supports for this research, and Clovis Sukam for his critical reading for this manuscript. The author also thanks Glenn Fei, Biao Wu, Zunwei Du, Wallace Law, and Kaijie Cui for many valuable discussions.

REFERENCES

[1] Ankarath, N., Ghost, T.P., Mehta, K.J., Alkafaji, Y. A. (2010), Understanding IFRS Fundamentals, John Wiley & Sons, Inc.

[2] Board of Governors of the Federal Reserve System (2016). Comprehensive Capital Analysis and Review 2016 Summary and Instructions, January 2016.

[3] Derbali, A., Hallara, S. (2013), Analysis of default probability: A comparative theoretical approach between the Credit Portfolio View model and the CreditRisk+ model, International Journal of Business Management & Research, Vol. 3 (1), 157-170

[4] Diaz, D., Gemmill, G. (2002), A systematic comparison of two approaches to measuring credit risk: CreditMetrics versus CreditRisk+, Trans 27th ICA

[5] Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. Journal of Financial Intermediation12, pp.199-232. doi:10.1016/S1042-9573(03)00040-8

[6] Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance, Volume 29 (2), 449-470 DOI: 10.1111/j.1540-6261.1974.tb03058.x

[7] Miu, P., Ozdemir, B. (2009). Stress testing probability of default and rating migration rate with respect to Basel II requirements, Journal of Risk Model Validation, Vol. 3 (4) Winter 2009

[8] Nystrom, K., Skoglund, J. (2006), A credit risk model for large dimensional portfolios with application to economic capital, Journal of banking & finance, Vol. 30, 2163-2197 DOI:10.1016/j.jbankfin.2005.05.024

[9] Rosen, D., Saunders, D. (2009). Analytical methods for hedging systematic credit risk with linear factor portfolios. Journal of Economic Dynamics & Control, 33 (2009), 37-52 http://www.r2-financial.com/wp-content/uploads/2010/07/LinearFactor.pdf

[10] Vasicek, O. (2002). Loan portfolio value. RISK, December 2002, 160 - 162.

[11] Wei, J. Z. (2003), A multi-factor, credit migration model for sovereign and corporate debts, Journal of international money and finance, Vol. 22, 709-735 DOI:10.1016/S0261-5606(03)00052-4

[12] Wolfinger, R. (2008). Fitting Nonlinear Mixed Models with the New NLMIXED Procedure. SAS Institute Inc.

[13] Yang, B. H. , Zunwei Du (2016). Rating Transition Probability Models and CCAR Stress Testing, Journal of Risk Model Validation 10 (3), 2016, 1-19 DOI: 10.21314/JRMV.2016.155

Appendix

Proof of Proposition 3.1. We show only (3.7B), the proof for (3.7A) is similar. For simplicity, we write F_i and $\tilde{p}_i$, respectively, for $F_i(x)$ and $\tilde{p}_i(x)$. The marginal probability for the event $\{R = i \mid x\}$ is $(1 - F_{i-1}(x))\tilde{p}_i(x)$. Thus the probability for observing a frequency n_i for the i^{th} rank value is:

$$(1 - F_{i-1})^{n_i}\, \tilde{p}_i^{\,n_i}$$

up to a multiplicative factor given by the binomial coefficient. Consequently, the probability observing frequencies $\{n_i\}_{i=1,2,\ldots,k}$ with n_i for the i^{th} rank value is:

$$\Delta = \tilde{p}_1^{\,n_1}\tilde{p}_2^{\,n_2}\ldots\tilde{p}_k^{\,n_k}(1 - F_1)^{n_2}(1 - F_2)^{n_3}\ldots(1 - F_{k-1})^{n_k} \qquad \text{(A.1)}$$

up to a constant factor given by some binomial coefficients. By (2.1D) of Proposition 2.1, we have:

$$\begin{aligned}
&(1 - F_1)^{n_2}(1 - F_2)^{n_3}\ldots(1 - F_{k-1})^{n_k} \\
&= (1 - \tilde{p}_1)^{n_2}[(1 - \tilde{p}_1)(1 - \tilde{p}_2)]^{n_3}\ldots[(1 - \tilde{p}_1)(1 - \tilde{p}_2)\ldots(1 - \tilde{p}_{k-1})]^{n_k} \\
&= (1 - \tilde{p}_1)^{n_2 + n_3 + \ldots + n_k}(1 - \tilde{p}_2)^{n_3 + n_4 + \ldots + n_k}\ldots(1 - \tilde{p}_{k-1})^{n_k}
\end{aligned}$$

$$= (1-\tilde{p}_1)^{s_1-n_1}(1-\tilde{p}_2)^{s_2-n_2}\ldots(1-\tilde{p}_{k-1})^{s_{k-1}-n_{k-1}} \qquad \text{(A.2)}$$

The equation (A.2) follows from (3.5B). Thus, by (A.1), the corresponding log-likelihood is:

$$\begin{aligned}\log(\Delta) &= [n_1\log(\tilde{p}_1)+(s_1-n_1)\log(1-\tilde{p}_1)]+[n_2\log(\tilde{p}_2)+(s_2-n_2)\log(1-\tilde{p}_2)]+\ldots+[n_k\log(\tilde{p}_k)] \\ &= [n_1\log(\tilde{p}_1)+(s_1-n_1)\log(1-\tilde{p}_1)]+[n_2\log(\tilde{p}_2)+(s_2-n_{n2})\log(1-\tilde{p}_2)]+\ldots \\ &\quad +[n_k\log(\tilde{p}_k)+(s_k-n_k)\log(1-\tilde{p}_k)] \qquad \text{(A.3)} \\ &= L_1(x)+L_2(x)+\ldots+L_k(x)\end{aligned}$$

where the equation (A.3) follows form the fact that $(s_k-n_k)=0.$ □

Proof of Proposition 3.2. It is well-known that the standard normal or logistic cumulative distribution is log concave. Also, if $f(x)$ is log concave, then so is $f(Az+b),$ where $Az+b: R^m \to R^1$ is any affine transformation from the *m*-dimensional Euclidean space to the one-dimensional Euclidean space. Therefore both the cumulative distributions $\Phi(x)$ and $\Phi(-x)$ are log concave. For Proposition 3.2 (a), the concavity of (3.7A) and (3.7B) follows from the fact that the sum of concave functions is again concave. For Proposition 3.2 (b), the concavity of (3.7A) and (3.7B) as a function of a-parameters is also immediate.

For Proposition 3.2 (b) and the concavity of (3.7A) and (3.7B), as a function of the r-parameters $\{r_i\}$, recall that $\tilde{p}_i(x)$ in (3.4) is given by

$$\tilde{p}_i(x) = \Phi(c_i\sqrt{1+(r_i v)^2}+r_i(a_1x_1+a_2x_2+\ldots+a_mx_m-u))$$

It suffices to show that the 2nd derivative of the function

$$L(r) = \log[\Phi(b\sqrt{1+r^2}+ra)] \qquad \text{(A.4)}$$

is non-positive for any constants a and $b.$ This is because either $\log(\tilde{p}_i(x)])$ or $\log(1-\tilde{p}_i(x)])$ will have the form of (A.4) after some appropriate scaling transformations. The 2nd derivative $d^2[L(r)]/dr^2$ is given by:

$$\begin{aligned}&(br/\sqrt{1+r^2}+a)^2\{-[\phi(b\sqrt{1+r^2}+ra)]^2/[\Phi(b\sqrt{1+r^2}+ra)]^2+\phi'(b\sqrt{1+r^2}+ra)/\Phi(b\sqrt{1+r^2}+ra)\} \\ &\qquad +\phi(b\sqrt{1+r^2}+ra)(b)(1+r^2)^{-3/2}/\Phi(b\sqrt{1+r^2}+ra) \\ &= I+II \qquad \text{(A.5)}\end{aligned}$$

where ϕ and ϕ' denote the 1st and 2nd derivatives of Φ. Because the factor in the 1st summand of (A.5)

$$\{-[\phi(b\sqrt{1+r^2}+ra)]^2/[\Phi(b\sqrt{1+r^2}+ra)]^2+\phi'(b\sqrt{1+r^2}+ra)/\Phi(b\sqrt{1+r^2}+ra)\}$$

corresponds to the 2nd derivative of $\log\Phi(z)$ (with respect to $z=b\sqrt{1+r^2}+ra),$ it is non-positive. Thus the 1st summand in (A.5) is non-positive. The 2nd summand in (A.5) is non-positive if $b \le 0.$ For the case $b>0,$ we can change b back to the negative case using the function $F(x)=\Phi(-x)$ and repeat the same discussion to have non-positivity of the 2nd derivative of (A.4).
□