Journal of Risk Model Validation **11**(3), 1–18
DOI: 10.21314/JRMV.2017.181

Research Paper

Forward ordinal probability models for point-in-time probability of default term structure: methodologies and implementations for IFRS 9 expected credit loss estimation and CCAR stress testing

Bill Huajian Yang

Enterprise Stress Testing, Royal Bank of Canada, 155 Wellington Street West, Toronto M5V 3H6, Canada; email: bill.yang@rbc.com

(Received January 31, 2017; revised June 23, 2017; accepted June 26, 2017)

ABSTRACT

Common ordinal models, including the ordered logit model and the continuation ratio model, are formulated by a common score (ie, a linear combination of given explanatory variables) plus rank-specific intercepts. Sensitivity to the common score is generally not differentiated between rank outcomes. We propose an ordinal model based on forward ordinal probabilities for rank outcomes. In addition to the common score and intercepts, the forward ordinal probabilities are formulated by the rank- and rating-specific sensitivity (for a risk-rated portfolio). This rank-specific sensitivity allows a risk rating to respond to its migrations to default, downgrade, stay and upgrade accordingly. A parameter estimation approach based on maximum likelihood for observing rank-outcome frequencies is proposed. Applications of the proposed model include modeling rating migration probability for point-in-time probability of default term structure for International Financial Reporting Standard 9 expected credit loss estimation and Comprehensive Capital Analysis and Review stress testing. Unlike the rating transition model based on the Merton model, which allows only one sensitivity

Print ISSN 1753-9579 | Online ISSN 1753-9587

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 to include entity-specific risk drivers (eg, the downgrade history or credit quality changes for an entity in the previous two quarters can be included). No additional estimation of the asset correlation is required. As an example, the proposed model, benchmarked with the rating transition model based on the Merton model, is used to estimate the probability of default term structure for a commercial portfolio, where for each rating the sensitivities are differentiated between migrations to default, downgrade, stay and upgrade. Our 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 $\{1, 2, \ldots, k\}$. The forward (respectively, backward) 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 (respectively, not larger) than i.

Common ordinal models, as reviewed in Section 2, include the ordered logit model (ie, 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 (ie, a linear combination of explanatory variables) together with a rank-specific intercept. For a continuation ratio model, the forward or backward ordinal probabilities for rank outcomes are also modeled by a common score with rank-specific intercept. Sensitivities to the common score are 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 than to lower risk ratings during downturns. Risk sensitivity is not generally uniform between risk ratings or between outcome ranks.

We propose an ordinal model based on forward ordinal probability (model (3.2) or (3.4); see Section 3). The forward ordinal probabilities are formulated by a common score plus rank-specific sensitivity and intercept. We propose an algorithm for parameter estimation based on maximum likelihood for observing rank-outcome frequencies. The model can be implemented easily by a modeler using, for example, the SAS software procedure PROC NLMIXED (Wolfinger 2008).

Applications of the proposed model include

(a) modeling the rating migration probability for Comprehensive Capital Analysis and Review (CCAR) stress testing (Board of Governors of the Federal Reserve System 2016) and the point-in-time probability of default (PD) term structure for International Financial Reporting Standard 9 (IRFS 9) expected credit loss estimation (Ankarath *et al* 2010);

(b) estimation of the PD for a low default portfolio, and shadow rating modeling.

The modeling of state transition probabilities dates back to the original CreditMetrics, CreditPortfolioView and CreditRisk+ credit portfolio approaches (Derbali and Hallara 2013; Diaz and Gemmill 2002), and contributions by researchers including Nyström and Skoglund (2006) and Wei (2003). The point-in-time rating transition probability model based on the Merton model (Gordy 2003; Merton 1974; Miu and Ozdemir 2009; Vasicek 2002), which is formulated by a common credit index, was proposed by Miu and Ozdemir (2009), and extended by Yang (2016) to facilitate rating-level sensitivity for CCAR stress testing and IFRS 9 expected credit loss estimation.

Our proposed ordinal model, formulated by a common score plus an 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. Monotonicity can be imposed for the sensitivity parameters between initial ratings for each type of migration. Under this model structure, risk for an entity is driven by the common score (as a dynamic) plus the sensitivity in responding to a scenario. Unlike the rating transition model (Yang 2016) based on the Merton model framework, which allows only one sensitivity parameter for all outcomes for a rating and uses only systematic risk drivers, our model can include entity-specific risk drivers and allows for rank-specific sensitivity. No additional estimation for asset correlation is required. Further, the loglikelihood based on the forward PD given by a cumulative distribution 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 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 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 loglikelihood function and its concavity. A heuristic hard expectation maximization algorithm for parameter estimation is proposed in Section 4. The model is validated and used in Section 5 to estimate the point-in-time PD structure for a commercial portfolio, where for each rating the sensitivities are differentiated between migrations to default, downgrade, stay and upgrade. The model is benchmarked using a rating transition model based on the Merton model framework. Concluding remarks are given in Section 6.

2 A REVIEW OF ORDINAL REGRESSION MODELS

In this section we review two common ordinal models: ordinal regression and the continuation ratio model.

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

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

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

$$\tilde{p}_i(x) = P(R = i \mid x, R \geqslant i), \qquad \tilde{p}_{\mathrm{b}i}(x) = P(R = i \mid x, R \leqslant i).$$

Remark 2.1 We can always model the backward ordinal probability via the forward ordinal probability model: we simply reverse the order of the ordinal outcomes and reindex the resulting forward ordinal probability $\tilde{p}_i(x)$ by replacing i with $(k + 1 - i)$. Therefore, we focus only on the forward ordinal probability model; all discussions for this model apply naturally to the backward ordinal model by an appropriate reversion for the outcome order and the index of the forward probability.

Proposition 2.2 *The following equations hold:*

$$F_i(x) = p_1(x) + p_2(x) + \cdots + p_i(x), \tag{2.1a}$$

$$\tilde{p}_i(x) = \frac{p_i(x)}{1 - F_{i-1}(x)}, \tag{2.1b}$$

$$p_i(x) = F_i(x) - F_{i-1}(x) = [1 - F_{i-1}(x)]\tilde{p}_i(x), \tag{2.1c}$$

$$[1 - F_i(x)] = [1 - \tilde{p}_1(x)][1 - \tilde{p}_2(x)] \cdots [1 - \tilde{p}_i(x)]. \tag{2.1d}$$

Proof Equation (2.1 a) is immediate. Equation (2.1 b) follows from the Bayesian theorem, while (2.1 c) follows from (2.1 a) and (2.1 b). By (2.1 c), we have

$$1 - F_i(x) = [1 - F_{i-1}(x)] - p_i(x) = [1 - F_{i-1}(x)][1 - \tilde{p}_i(x)].$$

Thus, (2.1 d) follows by induction. □

For the largest rank outcome k, we have

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

Therefore, by Proposition 2.2, an ordinal model can be chosen to model one of the following components:

(i) the cumulative probabilities $\{F_i(x) \mid i = 1, 2, \ldots, k-1\}$;

(ii) the marginal probabilities $\{p_i(x) \mid i = 1, 2, \ldots, k-1\}$;

(iii) the forward ordinal probabilities $\{\tilde{p}_i(x) \mid i = 1, 2, \ldots, k-1\}$.

Marginal probabilities are subject to the following constraints:

$$p_1(x) + p_2(x) + \cdots + p_i(x) \leqslant 1, \qquad p_1(x) + p_2(x) + \cdots + p_k(x) = 1.$$

Therefore, modeling marginal probabilities individually introduces additional complexity. In general, we can choose to model either the cumulative probabilities or the forward ordinal probabilities, as reviewed and discussed below.

2.1 Ordinal regression models

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

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

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

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

$$F_1(x) \leqslant F_2(x) \leqslant \cdots \leqslant F_{k-1}(x). \tag{2.3}$$

This is guaranteed for the ordinal regression model by the constraint $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_{k-1}$ in (2.2). When modeling the cumulative probabilities, condition (2.3) implies the coefficients $a_1, a_2, \ldots, a_m$ in (2.2) must be the same for all rank outcomes $\{i = 1, 2, \ldots, k-1\}$, a limitation of 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 ith outcome, is

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

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

$$\begin{aligned} \log\left[\frac{P(R \leqslant i \mid x)}{P(R > i \mid x)}\right] &= b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m \\ \implies F_i(x) &= P(R \leqslant i \mid x) \\ &= \frac{1}{1 + \exp(-b_i - a_1x_1 - a_2x_2 - \cdots - a_mx_m)} \\ &= F(b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m), \end{aligned}$$

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, ie, 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 (ie, the logit and probit functions).

2.2 Forward/backward continuation ratio model

Recall that the logit function is defined as $\operatorname{logit}(p) = \log[p/(1-p)]$ for $0 < p < 1$. Given scenario x and rank-outcome value i, the forward and backward logistic continuation ratio models, respectively, are formulated as

$$\operatorname{logit}\left[\frac{P(R = i \mid x)}{P(R \geqslant i \mid x)}\right] = b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m, \tag{2.4a}$$

$$\operatorname{logit}\left[\frac{P(R = i \mid x)}{P(R \leqslant i \mid x)}\right] = b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m. \tag{2.4b}$$

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

$$\begin{aligned} \tilde{p}_i(x) &= \frac{1}{1 + \exp(b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m)} \\ &= \Phi(b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m), \end{aligned} \tag{2.5}$$

where Φ denotes standard logistic cumulative distribution. This means the logistic forward continuation ratio model is formulated by the forward ordinal probabilities for rank outcomes, with the inverse of the standard logistic cumulative distribution, ie, the logit function, as the link function. The probit continuation ratio model can be formulated similarly using the inverse of the standard normal cumulative distribution (ie, the probit function) as the link function.

3 THE PROPOSED FORWARD ORDINAL MODEL

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

3.1 The mathematical setup

We assume, given the rank outcome will not be less than i, ie, $R \geqslant i$, that there is a latent variable y_i given by

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

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

By appropriate scaling of 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, by (3.1), the forward ordinal probability $\tilde{p}_i(x)$ is

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

Let $c(x) = (a_1x_1 + a_2x_2 + \cdots + a_mx_m)$. We call $c(x)$ a common score and r_i the sensitivity for the rank value $i \leqslant k - 1$ with respect to $c(x)$. For IFRS 9 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 (respectively, decrease) in the norm of the parameter vector $(a_1, a_2, \dots, a_m)$ during parameter estimation can propagate to the sensitivity parameter vector $(r_1, r_2, \dots, r_{k-1})$ by scaling down (respectively, up). To prevent unnecessary disturbance of parameter estimation and ensure estimation convergence, the following constraints can be imposed:

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

In practice, the sign of the coefficient a_i is predetermined. For example, default risk increases as unemployment rate increases. We thus require the unemployment rate coefficient 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 the following linear constraint can be imposed:

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

Let $c(x) = (a_1x_1 + a_2x_2 + \cdots + a_mx_m)$. In the case when the variables $x_1, x_2, \ldots, x_m$ are common to all entities (eg, the macroeconomic variables), we obtain the following model, assuming normality for $c(x)$ with mean u and standard deviation v:

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

where c_i is the threshold value estimated directly by taking the inverse, Φ^{-1}, of 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 (Rosen and Saunders 2009) for the expectation with respect to s:

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

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

3.2 The loglikelihood function given the observed rank frequencies

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

Given a scenario $x = (x_1, x_2, \ldots, x_m)$, let n_i denote the corresponding observed frequency for the ith rank value. Let

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

Define

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

We focus on the conditional probability space given that the rank value of the outcome R is not less than i. The loglikelihood of observing frequency n_i for the ith rank value and frequency $s_i - n_i$ for rank values larger than i, given $x = (x_1, x_2, \ldots, x_m)$, 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 the model parameters of model (3.2) and (3.4), assuming the occurrence of the ith rank value is a binary event.

Let $L(x, i, i+h)$ denote the loglikelihood, over this probability space, of observing multiple frequencies $\{n_i, n_{i+1}, \dots, n_{i+h}\}$, for rank values $\{i, i+1, \dots, i+h\}$, and the frequency $s_{i+h+1} = n_k + n_{k-1} + \cdots + n_{i+h+1}$ for rank values larger than $i+h$. We have the following proposition.

PROPOSITION 3.1 *The equations*

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

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

hold, up to a summand given by the logarithms of some binomial coefficients (independent of the parameters in models (3.2) and (3.4)).

PROOF 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 of the event $\{R = i \mid x\}$ is $(1 - F_{i-1}(x))\tilde{p}_i(x)$. Thus, the probability of observing a frequency n_i for the ith 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 of observing frequencies $\{n_i\}_{i=1,2,\dots,k}$ with n_i for the ith rank value is

$$\Delta = \tilde{p}_1^{n_1} \tilde{p}_2^{n_2} \cdots \tilde{p}_k^{n_k} (1 - F_1)^{n_2} (1 - F_2)^{n_3} \cdots (1 - F_{k-1})^{n_k} \tag{3.8}$$

up to a constant factor given by some binomial coefficients. By (2.1d), we have

$$\begin{aligned}(1 - F_1)^{n_2}&(1 - F_2)^{n_3} \cdots (1 - F_{k-1})^{n_k} \\ &= (1 - \tilde{p}_1)^{n_2}[(1 - \tilde{p}_1)(1 - \tilde{p}_2)]^{n_3} \cdots [(1 - \tilde{p}_1)(1 - \tilde{p}_2) \cdots (1 - \tilde{p}_{k-1})]^{n_k} \\ &= (1 - \tilde{p}_1)^{n_2+n_3+\cdots+n_k}(1 - \tilde{p}_2)^{n_3+n_4+\cdots+n_k} \cdots (1 - \tilde{p}_{k-1})^{n_k} \\ &= (1 - \tilde{p}_1)^{s_1-n_1}(1 - \tilde{p}_2)^{s_2-n_2} \cdots (1 - \tilde{p}_{k-1})^{s_{k-1}-n_{k-1}}.\end{aligned} \tag{3.9}$$

This follows from (3.5b). Thus, by (3.8), we have the corresponding loglikelihood:

$$\begin{aligned}\log \Delta &= [n_1 \log(\tilde{p}_1) + (s_1 - n_1)\log(1 - \tilde{p}_1)] \\ &\qquad + [n_2 \log(\tilde{p}_2) + (s_2 - n_2)\log(1 - \tilde{p}_2)] + \cdots + [n_k \log(\tilde{p}_k)] \\ &= [n_1 \log(\tilde{p}_1) + (s_1 - n_1)\log(1 - \tilde{p}_1)] \\ &\qquad + [n_2 \log(\tilde{p}_2) + (s_2 - n_{n2})\log(1 - \tilde{p}_2)] + \cdots \\ &\qquad + [n_k \log(\tilde{p}_k) + (s_k - n_k)\log(1 - \tilde{p}_k)] \\ &= L_1(x) + L_2(x) + \cdots + L_k(x),\end{aligned} \tag{3.10}$$

where the second equality follows from the fact that $(s_k - n_k) = 0$. □

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 finding the maximum likelihood estimate.

PROPOSITION 3.2 *The loglikelihood function (3.7a) or (3.7b), with Φ the standard normal cumulative probability distribution, is concave in the following two cases:*

(a) *as a function of the r-parameters $\{r_i\}$, or the b-parameters $\{b_j\}$, and the a-parameters $\{a_1, a_2, \dots, a_m\}$, where $\tilde{p}_i(x)$ is given by (3.2);*

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

PROOF It is well known that the standard normal cumulative distribution is log concave. Also, if $f(x)$ is log concave, then so is $f(Az + b)$, where $Az + b \colon \mathbb{R}^m \to \mathbb{R}^1$ 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 parameters $\{a_1, a_2, \dots, a_m\}$ 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

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

It suffices to show that the second derivative of the function

$$L(r) = \log[\Phi(b\sqrt{1 + r^2} + ra)] \tag{3.11}$$

is nonpositive 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 (3.11) after some appropriate scaling transformations. The second derivative $\mathrm{d}^2[L(r)]/\mathrm{d}r^2$ is given by

$$\begin{aligned}\left(\frac{br}{\sqrt{1 + r^2}} + a\right)^2 &\left\{-\frac{[\varphi(b\sqrt{1 + r^2} + ra)]^2}{[\Phi(b\sqrt{1 + r^2} + ra)]^2} + \frac{\varphi'(b\sqrt{1 + r^2} + ra)}{\Phi(b\sqrt{1 + r^2} + ra)}\right\} \\ &+ \frac{\varphi(b\sqrt{1 + r^2} + ra)(b)(1 + r^2)^{-3/2}}{\Phi(b\sqrt{1 + r^2} + ra)} = \mathrm{I} + \mathrm{II},\end{aligned} \tag{3.12}$$

where φ and φ' denote the first and second derivatives of Φ. Because the factor in the first summand of (3.12),

$$\left\{-\frac{[\varphi(b\sqrt{1 + r^2} + ra)]^2}{[\Phi(b\sqrt{1 + r^2} + ra)]^2} + \frac{\varphi'(b\sqrt{1 + r^2} + ra)}{\Phi(b\sqrt{1 + r^2} + ra)}\right\},$$

corresponds to the second derivative of $\log\Phi(z)$ (with respect to $z = b\sqrt{1 + r^2} + ra$), it is nonpositive. Thus, the first summand in (3.12) is nonpositive. The second summand in (3.12) is nonpositive if $b \leqslant 0$. For the $b > 0$ case, we can change b back to the negative case using the function $F(x) = \Phi(-x)$ and repeat the same discussion to obtain nonpositivity of the second derivative of (3.11). □

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 loglikelihood for observing rank-outcome frequencies. This generic algorithm works for one forward ordinal model. For modeling rating migration for a risk-rated portfolio, multiple forward ordinal models are required: one for each of the nondefault initial risk ratings (see Section 5 for the model formulation and the adapted algorithm for parameter fitting).

4.1 Estimation of parameters for model (3.2)

The algorithm proposed is essentially a heuristic hard expectation maximization algorithm.

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

Step 1: assume that the parameters $\{a_1, a_2, \dots, a_m\}$ and $\{b_1, b_2, \dots, b_{k-1}\}$ are given. Estimate the sensitivities $\{r_1, r_2, \dots, r_{k-1}\}$ by maximizing the loglikelihood of (3.7 b).

Recall that, by Proposition 3.2(a), global maximum estimates are guaranteed.

Step 2: assume that the sensitivities $\{r_1, r_2, \dots, r_{k-1}\}$ are given. Estimate the parameters $\{a_1, a_2, \dots, a_m\}$ and $\{b_1, b_2, \dots, b_{k-1}\}$ by maximizing the loglikelihood of (3.7 b).

Global maximum estimates are granted by Proposition 3.2(a). Rescale the a-parameter estimates by a scalar $\rho > 0$ to make $(a_1, a_2, \dots, a_m)$ a unit vector, and then scale the vector $(r_1, r_2, \dots, r_{k-1})$ by $1/\rho$.

Step 3: iterate the above three steps until a convergence is reached.

Steps 1 and 2 are repeated until convergence is reached, ie, the maximum deviation for all parameter estimates for $\{b_1, b_2, \dots, b_{k-1}\}$, $\{a_1, a_2, \cdots, a_m\}$ and $\{r_1, r_2, \dots, r_{k-1}\}$, in two consecutive iterations, is less than 10^{-4}.

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

4.2 Estimation of parameters for model (3.4)

For model (3.4), we follow the steps above to fit for the coefficients $\{a_1, a_2, \cdots, a_m\}$ for common score $c(x) = (a_1x_1 + a_2x_2 + \cdots + a_mx_m)$. When this common score is known, we estimate $\{r_1, r_2, \ldots, r_{k-1}\}$ by maximizing (3.7b), with $\tilde{p}_i(x)$ given by (3.4). Global maximum estimates are guaranteed by Proposition 3.2(b).

5 AN EMPIRICAL EXAMPLE: RATING MIGRATION PROBABILITY AND PROBABILITY OF DEFAULT 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, where a point-in-time PD term structure for IFRS 9 expected credit loss estimation and CCAR stress testing is derived.

The sample contains quarterly rating migration frequencies between 2006 Q3 and 2016 Q4 for a commercial portfolio, created synthetically by scrambling the default rate using an appropriate scaling. There are twenty-one risk ratings, with R_{21} 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 with the backward ordinal model, starting with a 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 of the index of the resulting forward ordinal probability.

The backward ordinal model is benchmarked using the rating transition model based on the Merton model framework proposed by Yang (2016). Additional benchmark comments for SAS ordinal regression using SAS PROC LOGISTIC are given at the end of this section.

5.1 The backward ordinal and benchmark models for IFRS 9 expected credit loss estimation and CCAR stress testing

5.1.1 Formulation of the models

Backward ordinal model for rating migration probability.

Given a nondefault initial risk rating R_i at the beginning of the quarter, there are twenty-one possible ordinal outcomes at the end of the quarter: an entity can migrate to a default rating or any of the other twenty ratings. Given a scenario $x = (x_1, x_2, \ldots, x_m)$, 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 higher 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 $\tilde{p}_{ij}(x)$ using models (3.2) and (3.4), respectively, as follows:

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

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

We assume that, for each initial rating R_i, the sensitivity parameters r_{ij} are the same for rank-outcome values j when $i < j < 21$ (downgrade) or $1 \leqslant j < i$ (upgrade). Denote the downgrade sensitivity by $r_{i\mathrm{d}}$ and the upgrade sensitivity by $r_{i\mathrm{u}}$. Let $r_{i\mathrm{df}}$ and $r_{i\mathrm{s}}$ denote the sensitivities for outcome cases $j = 21$ (default) and $j = i$ (stay), respectively. Then (5.1 a) and (5.1 b) reduce to

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

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

where $r_i = r_{i\mathrm{df}}, r_{i\mathrm{d}}, r_{i\mathrm{s}}, r_{i\mathrm{u}}$, 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 $F_{ij}(x) = p_{i21}(x) + p_{i20}(x) + \cdots + p_{i22-j}(x)$ is the cumulative probability. The constraint (3.3 a) (respectively, (3.3 b)) is imposed for the proposed backward ordinal model (5.2 a) (respectively, (5.2 b)).

Rating transition model under the Merton model framework

The point-in-time rating transition probability model, based on the Merton framework, was proposed by Miu and Ozdemir (2009), and extended by Yang (2016) to facilitate rating-level sensitivity for CCAR stress testing and IFRS 9 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 $x = (x_1, x_2, \ldots, x_m)$. Let Φ denote the standard normal cumulative distribution. Under the Merton model framework (Gordy 2003; Merton 1974; Miu and Ozdemir 2009; Vasicek 2002), it can be shown (Yang 2016) that

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

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 R_i

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

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

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

$$\tilde{r}_i = \frac{r_i\lambda}{\sqrt{1 + r_i^2(1-\lambda^2)}}, \quad 0 \leqslant \lambda \leqslant 1, \tag{5.5}$$

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 (Yang 2016).

Remark 5.1 We can choose to fit for $\{a_1, a_2, \dots, a_m\}$ without constraint (5.5). The unconstrained result is always better than the constrained one in the sense of a higher likelihood value.

5.1.2 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 $\{a_1, a_2, \dots, a_m\}$ by using maximum likelihood to observe rating-level default frequencies, with default probability $p_i(x)$ for rating R_i given by (5.4) without constraint (5.5). This can be done via Steps 1–3 in Section 4.

(2) When credit index $\mathrm{ci}(x) = \tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + \cdots + \tilde{a}_m\tilde{x}_m$ has been determined, we need to fit only for the risk sensitivity parameters $\{r_i\}$ for model (5.3), and $\{r_{i\mathrm{df}}, r_{i\mathrm{d}}, r_{i\mathrm{s}}, r_{i\mathrm{u}}\}$ for model (5.2b), for all risk ratings. For model (5.3), we can choose to fit for $\{r_i\}$ either separately for each rating R_i, or as a combination of all ratings, using the appropriate likelihood function (3.7b) for all rating migration frequencies, or (3.7a) for downgrade or default frequencies only. The corresponding loglikelihood function is concave by Proposition 3.2(b). For model (5.2b), we fit for each of the four groups $\{r_{i\mathrm{df}}\}$, $\{r_{i\mathrm{d}}\}$, $\{r_{i\mathrm{s}}\}$ and $\{r_{i\mathrm{u}}\}$ separately, using the appropriate likelihood function (3.7a) for the corresponding migration frequency.

TABLE 1 Macro coefficients.

Model	v_1	v_2	p_1	p_2
BORD	0.3975	0.6025	0.0247	<0.0001
RTGM	0.3975	0.6025	0.0247	<0.0001

In general, monotonicity is imposed for sensitivity between ratings; specifically, we require that $\{r_i\}$, $\{r_{i\mathrm{df}}\}$ and $\{r_{i\mathrm{d}}\}$ are nondecreasing and that $\{r_{i\mathrm{s}}\}$ and $\{r_{i\mathrm{u}}\}$ are nonincreasing for a higher risk rating.

5.2 Validation results

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

- BORD: the backward ordinal model (5.2 b).
- RTGM: the rating migration model based on the Merton model framework (5.3).

Both models use the same variables, provided by the US Federal Reserve:

- three-month treasury bill interest rate (v_1);
- unemployment rate (v_2).

The macro coefficients for credit index $\mathrm{ci}(x) = a_1x_1 + a_2x_2 + \cdots + a_mx_m$ are fitted as described in Section 5.1 in the same way for both the BORD and RTGM, so both models have the same macro coefficient estimates. Table 1 records the estimates for these two variable coefficients, with the variable p-values p_1 and p_2.

For the BORD the sensitivity parameter estimates are reported as in Table 2 for twenty nondefault ratings for default, downgrade and stay, with monotonicity constraint being imposed. The sensitivity estimates for upgrade are all close to zero (reflecting the fact that the upgrade probability is slim in the stress period), and are not shown in the table. The RTGM estimates the sensitivity parameters by maximum likelihood for observing only the default frequency. Thus, for default it has the same sensitivity estimates as the BORD.

Table 3 shows the backtest performance for two R-squared-based models for predicting portfolio cumulative default rates for one, four, six, twelve and sixteen quarters.

The results show performance improves for the BORD when the sensitivities are differentiated between migrations to default, downgrade, stay and upgrade. This improvement is a trade-off with the addition of more sensitivity parameters.

TABLE 2 Sensitivity parameter estimates.

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

TABLE 3 R-squared values for predicting portfolio cumulative default rate.

	Quarters				
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

We end this section by commenting on an additional benchmark 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 the "by" statement is used for initial ratings, for each initial rating R_i, SAS fits an ordinal regression model of the form

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

for the cumulative probability for rank outcome $j < 21$. This model has redundant coefficients (depending on the rating index i) for such a short time series sample, causing an overfitting issue. More importantly, it is not formulated using a common score or 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 for each initial risk rating R_i, SAS fits an ordinal model of the form

$$F_{ij}(x) = \Phi(b_{ij} + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m)$$

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

$$(b_{i1}, b_{i2}, \ldots, b_{i20}) = (d_i + b_{11}, d_i + b_{12}, \ldots, d_i + b_{120}) \tag{5.6}$$

with constant d_i corresponding to the ith 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 overestimates PD for the high risk ratings R_{19} and R_{20}, and significantly underestimates the PD for other ratings. We do not recommend this model.

6 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 proposed an ordinal model based on forward ordinal probabilities. Under this model, forward ordinal probabilities are formulated by a common score plus rank- and rating-specific sensitivity. The rank-specific sensitivity allows a risk rating to respond to its own migration patterns to default, downgrade, stay and upgrade accordingly. Empirical results show our model is more robust than the rating transition model based on the Merton model framework. Unlike the latter model, which allows only one sensitivity parameter for all rank outcomes for a rating, and uses only systematic risk drivers, our proposed ordinal model differentiates sensitivities between outcomes and includes 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 useful tool for practitioners to estimate the point-in-time PD term structure for IFRS 9 expected credit loss estimation as well as multiperiod scenario loss projection for CCAR stress testing.

DECLARATION OF INTEREST

The author reports no conflicts of interest. The author alone was responsible for the content and writing of the paper. The views expressed in this paper are not necessarily those of Royal Bank of Canada or any of its affiliates.

ACKNOWLEDGEMENTS

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

REFERENCES

Ankarath, N., Ghosh, T. P., Mehta, K. J., and Alkafaji, Y. A. (2010). *Understanding IFRS Fundamentals: International Financial Reporting Standards*. Wiley.

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

Derbali, A., and 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 and Research* **3**(1), 157–170.

Diaz, D., and Gemmill, G. (2002). A systematic comparison of two approaches to measuring credit risk: CreditMetrics versus CreditRisk+. In *Transactions of the 27th International Congress of Actuaries*. International Actuarial Association, Ottawa, ON.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation* **12**(3), 199–232 (http://doi.org/dffh36).

Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* **29**(2), 449–470 (http://doi.org/bhm7).

Miu, P., and Ozdemir, B. (2009). Stress-testing probability of default and migration rate with respect to Basel II requirements. *The Journal of Risk Model Validation* **3**(4), 3–38.

Nyström, K., and Skoglund, J. (2006). A credit risk model for large dimensional portfolios with application to economic capital. *Journal of Banking and Finance* **30**(8), 2163–2197 (http://doi.org/djr5d8).

Rosen, D., and Saunders, D. (2009). Analytical methods for hedging systematic credit risk with linear factor portfolios. *Journal of Economic Dynamics and Control* **33**(1), 37–52 (http://doi.org/b6qjmj).

Vasicek, O. (2002). Loan portfolio value. *Risk* **15**(12), 160–162.

Wei, J. Z. (2003). A multi-factor, credit migration model for sovereign and corporate debts. *Journal of International Money and Finance* **22**(5), 709–735 (http://doi.org/bmknzg).

Wolfinger, R. D. (2008). Fitting nonlinear mixed models with the new NLMIXED procedure. Data Analysis Paper, SAS Institute, Inc.

Yang, B. H., and Du, Z. (2016). Rating transition probability models and CCAR stress testing. *The Journal of Risk Model Validation* **10**(3), 1–19 (http://doi.org/bvfb).