Modelling Credit Grade Migration in Large Portfolios using cumulative t-link transition models

Jonathan J. Forster*
University of Southampton, UK†

Matteo Buzzacchi
Barclays Capital, UK†

Agus Sudjianto
Wells Fargo, USA†

Risa Nagao
University of St Andrews, UK†

Abstract

For a credit portfolio, we are often interested in modelling the migration of accounts between credit grades over time. For a large retail portfolio, data on credit grade migration may be available only in the form of a series of (typically monthly) population transition matrices representing the gross flow of accounts between each pair of credit grades in the given time period. The challenge is to model the transition process on the basis of these aggregate flow matrices. Each row of an observed transition matrix represents a sample from an ordinal probability distribution. Following Malik and Thomas (2012), Feng et al (2008) and McNeil and Wendin (2006), we assume a cumulative link model for these ordinal distributions. Common choices of link function are based on the normal (probit link) or logistic distributions, but the fit to observed data can be poor. In this paper, we investigate the fit of alternative link specifications based on the t-distribution. Such distributions arise naturally when modelling data which arise through aggregating an inhomogeneous sample of obligors, by combining a simple structural-type model for credit migration at the obligor level, with a suitable mixing distribution to model the variability between obligors.

Keywords: Markov Processes, Cumulative link, Heavy-tailed, Logistic, Probit, Transition matrix

*Corresponding author. Address: Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. Tel: +44 (0)23 80595130. Email: J.J.Forster@soton.ac.uk.

†This article represents the views and analysis of the authors only and should not be taken to represent those of any of their current or former employers. Much of this work was carried out while the first three authors were colleagues, and the fourth author a project student, at Lloyds Banking Group. The authors are grateful for the helpful comments of two anonymous reviewers on an earlier version of this paper.

1 Introduction

1.1 Background

Credit ratings can be an invaluable tool for describing and modelling the default risk for obligors in a particular loan pool. A credit rating system is an ordinal classification reflecting the probability of default of a given obligor, with the highest rating representing lowest probability of default, ranging down to a lowest rating typically representing an obligor already in default. For corporate assets, various ratings agencies (for example Moody's or Standard and Poor's) provide credit ratings. For retail obligors, a bank typically uses its internal credit scoring systems for rating purposes. Our work in this area is motivated by the need to model credit grade transitions in retail portfolios. Hence, we assume that grades are only observed at fixed discrete time intervals, and that detailed obligor-level covariate information is not available. However, we observe the same behaviour in retail portfolios and in pooled corporate agency ratings, and it is the latter which we use to illustrate our proposed modelling approach.

One possible method for forecasting the evolution of default risk in a portfolio is to forecast the process by which individual credit grades (including default) migrate over time. The natural description of the process of credit grade migration between two time points is the transition matrix $P(t,u)$, with elements $p_{ij}(t,u)$, representing the probability of transition from grade i at time point t to grade j at time point u, that is

$$p_{ij}(t,u) = \text{Prob}(\text{grade at time } u = j |\text{ grade at time } t = i).$$

Here, we assume that i and j run from 1 (highest quality) to D with the final grade D representing default, and that, as described above, the grades are naturally ordinal with increasing grade number representing increasing closeness to default.

If $\pi(t) = \{\pi_1(t), \ldots, \pi_D(t)\}$ is the row vector containing the proportions of obligors in each of the credit grades at time t, then a forecast of the corresponding proportions $\pi(u)$ at time $u > t$ is given by

$$\pi(u) = \pi(t)P(t,u).$$

Hence, forecasting a future credit grade profile can be achieved by estimating the corresponding transition matrix between the present and the forecast horizon. Estimators are constructed by developing statistical models for transition matrices, and fitting them to observed data on historical transitions. In this paper, we focus on developing statistical models which fit historical transition data on portfolio credit grade distributions. Typically, these data are a series of portfolio transition matrices representing the gross flow of obligors between each pair of credit grades in the given time period. We do not consider here the question of modelling the complete transition process. That is, given an empirical matrix $X(t,u)$ of portfolio flow between times t and u, we consider the problem of estimating the corresponding transition matrix $P(t,u)$ which can be thought of as a smoothed (and normalised so that rows sum to one) version of $X(t,u)$. The matrices $X(t,u)$ and $P(t,u)$ are $D \times D$ matrices, with elements $x_{ij}(t,u)$ representing the observed number of transitions from state i to state j between times t and u, and $p_{ij}(t,u)$ the corresponding transition probability. As the default state D is an absorbing state, estimation of $P(t,u)$ simply requires estimation of the first $D-1$ rows of $P(t,u)$.

We model the transition process over time as a series of one-period models, rather than a single all-encompassing model for the panel of empirical transition matrices. For a complete forecasting model, it is necessary to augment these one-period models for individual historical transitions, in order to predict future transition dynamics. This is discussed briefly in Section 5 and is the subject of ongoing research. Henceforth, for clarity, the dependence of P, (p_{ij}) and X (x_{ij}) on t

and u, the beginning and end points of the period under consideration, is considered as implicit, and omitted from our notation.

1.2 Ordinal data models

Ordinal data can often be effectively modelled using a cumulative link model. A cumulative link model, for a collection of ordinal variables $\{Y_k\}$, can be written as

$$P(Y_k \leq j) = g(\alpha_j - \mu_k) \tag{1}$$

for some strictly increasing function g which can be interpreted as the distribution function of a latent continuously distributed variable. Then, $-\infty = \alpha_0, \alpha_1, \ldots, \alpha_D = \infty$, can be thought of as an increasing sequence of thresholds, defining the mapping between the underlying latent scale and the ordinal classes. The μ_k parameters are observation-specific, but are typically modelled using a parsimonious regression function. Common choices of g are based on the standard normal (ordinal probit model) or logistic distributions (proportional odds model). A visual illustration of the mapping between the threshold parameters $\alpha_1, \ldots, \alpha_{D-1}$, the mean parameter μ and the class probabilities is given in Figure 1 (for the case where $D = 10$).

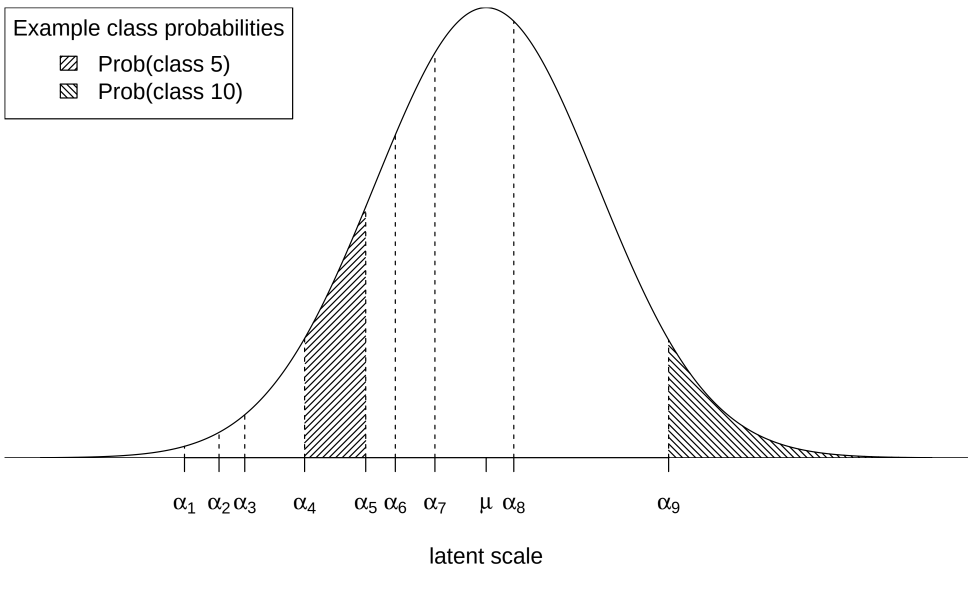

Figure 1: Mapping between threshold and mean parameters and class probabilities for a cumulative link model

For ordinal transition matrix modelling, each row represents the ordinal outcome distribution for a different originating class. If we treat each row X_i of our data matrix X (data on transitions from a single originating class i) as arising from observations of independent and identically distributed ordinal random variables, then these observations share a common value of μ_k in (1) denoted μ_i to acknowledge its dependence on the originating class i. Similarly, these ordinal outcomes share common threshold parameters, $\alpha_{i1}, \ldots, \alpha_{i\,D-1}$. A general model is therefore

$$q_{ij} \equiv \sum_{k=1}^{j} p_{ik} = g(\alpha_{ij} - \mu_i) \tag{2}$$

so q_{ij} are the cumulative transition probabilities for row i, and we assume the same form of underlying latent distribution (link function g) for each row. In fact, this is not a restriction, as any transition matrix can be fitted exactly by (2) whatever the specification for g. However, the fidelity of a more parsimonious specification for $\alpha_{ij} - \mu_i$ in (2) will depend on the form of g. In this paper, we consider two simplifications of (2), the standard cumulative link specification, introduced in (1), which for a transition matrix can be written as

$$q_{ij} = g(\alpha_j - \mu_i) \tag{3}$$

and the scale-varying cumulative link model

$$q_{ij} = g\left(\frac{\alpha_j - \mu_i}{\sigma_i}\right). \tag{4}$$

The standard cumulative link model (3) assumes a set of common thresholds, α, with the difference between rows of the transition matrix being represented by a shift (mean-change) in the distribution of the underlying latent variable. The scale-varying model allows for a 'shift and stretch' with the latent distribution differing between rows in both location (μ) and dispersion (σ). Note that (4) gives the same probabilities under addition of the same constant to α and μ, or multiplication of α, μ and σ by the same (positive) constant. Hence, to identify the parameters, we set $\mu_1 = 0$ and $\sigma_1 = 1$.

1.3 The link function

The natural (canonical) choice of link function is the logistic link

$$g(x) = \frac{\exp x}{1 + \exp x}$$

which is the distribution function of the standard logistic distribution. A cumulative logistic model is often referred to as a proportional odds model (McCullagh, 1980) as, for any $j \in \{1, \ldots, D-1\}$, from (1),

$$\frac{P(Y_k \le j)}{P(Y_k > j)} = \exp(\mu_\ell - \mu_k)\frac{P(Y_\ell \le j)}{P(Y_\ell > j)}.$$

Malik and Thomas (2012) use a cumulative logit model for credit grade transition dynamics in a retail portfolio.

Alternatively, one might choose the ordinal probit model, where $g = \Phi$, the standard normal distribution function. This model is intuitively attractive as it corresponds to a discretely observed version of an underlying structural Gaussian model for a latent asset-value variable. Suppose that an obligor's 'asset-value', which represents their ability to repay the loan follows a continuous-time geometric Brownian motion with drift μ and volatility σ^2, as in the classic Merton model (Merton, 1974). Then, the negative-log-asset value Z_{t+1}, given the value at the previous time $Z_t = z$ is normally distributed with mean $z + \lambda \equiv z + \sigma^2/2 - \mu$ and variance σ^2. Default occurs when the asset-value decreases below a threshold, or equivalently the negative-log-asset value exceeds an equivalent threshold. Credit grades other than default can also be thought of as being defined by thresholding asset-value. Hence an ordinal probit model (3) with $g = \Phi$, for credit grade transitions can be thought of as an approximation to a structural asset-value model where the α_j define the credit grade thresholds on the negative-log asset value scale, and $\mu_i = z + \lambda$ for a 'representative' value of $z \in (\alpha_{i-1}, \alpha_i]$. The approximation arises because only the credit grade is observed, so we have imperfect conditioning information $z \in (\alpha_{i-1}, \alpha_i]$, rather than observing the exact asset value. Conditioning on $z \in (\alpha_{i-1}, \alpha_i]$ in the Merton model results in a distribution for Z_{t+1} which is not exactly normal.

1.4 Previous work

The use of ordinal probit models for credit grades was investigated for corporate credit ratings by Nickell et al (2000) and for sovereign credit ratings by Hu et al (2002). In both cases, a form of model (3) is used, but where μ_i is allowed to depend on obligor-level characteristics. Feng et al (2008) propose a scale-varying ordinal probit model for corporate grade transitions, similar to (4). The logistic link was proposed by McNeil and Wendin (2006) for corporate grade migration and by Malik and Thomas (2012) in the retail setting. Again, a form of model (3) is proposed, but Malik and Thomas (2012) find dependence of μ_i on both an obligor-level characteristic (time since origination of the loan) and on the previous (time $t-1$) grade, indicating non-Markovian credit grade dynamics.

In all these approaches, parameters of the ordinal probit models are allowed to vary over time, either as a function of observable time-dependent covariates, or as a serially correlated stochastic process. As we have described above, the mean parameter of the underlying latent (normal or logistic) variable is allowed to vary between obligors, but all these approaches assume homogeneous dispersion across the obligor pool at a given time.

We also note here that other approaches are possible for modelling corporate grade transitions where more detailed information may be available. For example, Lando and Skødeberg (2002) propose a continuous-time transition model, based on a transition intensity matrix, while Chan et al (2012) use equity values to calibrate a structural transition model.

2 Exploratory analysis

We fit the model (4) to a credit card portfolio with $D = 12$, for both logistic and probit links. We also fitted the same models to an artificially created portfolio of corporate assets constructed from Moody's Default and Recovery Database (DRD), where $D = 8$ and the non-default grades are Aaa, Aa, A Baa, Ba, B and C. Finer numerical distinctions, such as Baa1 etc are ignored for the purposes of this analysis, and we also merge Caa, Ca and C into a single grade (C) due to smaller numbers of observations within these grades. Here, we present results from the latter analysis; the results of the retail analysis are very similar in character. One difference is that the retail portfolio is very much more mobile and larger (by several orders of magnitude), so almost all transitions are observed in a month. For the smaller corporate portfolio we therefore use a longer time period (a five-year transition matrix, over 2006-11) to exhibit a similar range of observed transitions. It is the analysis of this matrix we describe initially, and use to motivate further developments.

One might perform formal statistical goodness-of-fit analysis, either by conventional significance tests, or by comparing a fitted model with the saturated (unstructured) model for the transition matrix under study using the Akaike Information Criterion (AIC) or Bayes Information criterion

(BIC). Initially, however, we here adopt a less formal, graphical, assessment. From (4), we have

$$g^{-1}(q_{ij}) = \frac{1}{\sigma_i}\alpha_j - \frac{\mu_i}{\sigma_i}. \tag{5}$$

To assess the goodness-of-fit of (4) for a given link g, we plot, for each row i of the transition matrix, $g^{-1}(\hat{q}_{ij})$ against $\hat{\alpha}_j$ where

$$\hat{q}_{ij} = \frac{\sum_{k=1}^{j} x_{ik}}{\sum_{k=1}^{D} x_{ik}}$$

are the empirical cumulative transition probabilities and $\hat{\alpha}_j$ are the maximum likelihood estimates of the threshold parameters, obtained by maximising the log-likelihood

$$\ell(\alpha, \mu, \sigma) = \sum_{i,j} x_{ij} \log\left(g\left(\frac{\alpha_j - \mu_i}{\sigma_i}\right) - g\left(\frac{\alpha_{j-1} - \mu_i}{\sigma_i}\right)\right).$$

We use the R package `ordinal` for maximum likelihood estimation for model (4); see Christensen (2012) for details. If the model fits, then we should observe a series of approximately straight lines with gradient $1/\hat{\sigma}_i$ and intercept $-\hat{\mu}_i/\hat{\sigma}_i$. If the conventional cumulative link model (3) fits, then those lines will be parallel. Figure 2 presents this assessment for the probit and logistic link functions for the five year (2006-11) transition matrix, where we have collapsed the A grades into a single category, to make the visual representation clearer. It is immediately apparent that the straight lines (derived from the maximum likelihood estimates for μ and σ) do not fit the observed data well. The logistic is somewhat better than the probit, which together with the 'S'-shape of the empirical curves suggests that a heavier tailed link function might be more appropriate. In Figure 3, we present the equivalent graphical summary for a link based on the distribution function of a Student-t distribution with 2.65 degrees of freedom. (The value 2.65 is chosen to optimise the fit; see Section 3 for details.) The empirical curves are generally closer to straight lines indicating a superior fit, although we note that the gradients differ between the rows, so the scale-varying model (4) is required. One can also perform a formal goodness-of-fit test of the logistic, probit and t-link models. The p-values for log-likelihood ratio tests comparing these three models against the saturated alternative are 0.0007 (logistic) 0.0000 (probit) and 0.5124 (t-link). This indicates a failure to fit of logistic and probit models, but an adequate fit of the t-link model, suggesting that the main failure of the logistic and probit models is in the shape of the link function rather than in other aspects (such as the assumption of a common set of threshold parameters). In the next section we discuss a possible justification for the t-link and describe our approach for fitting a t-link model to a credit grade transition matrix.

3 The cumulative t-link model

Albert and Chib (1991) propose the use of a cumulative t-link model for ordinal data, where g in (4) is the distribution function of a standard Student-t distribution. This provides a family of link functions, depending on the degrees of freedom parameter ν of the t distribution. In the limit as ν tends to infinity, the t-link model converges to the probit model (as the t distribution converges to normality). Ntzoufras et al (2003) investigated the t-link model with unknown degrees of freedom ν. Applications of t-link modelling for ordinal data include the genetic analysis of Kizilkaya et al (2003).

For finite degrees of freedom, the t distribution is heavier tailed than the normal, admitting more extreme realisations. The left hand panel of Figure 2 indicates that a heavier-tailed link than probit is required for this transition matrix. The standard logistic distribution is heavier tailed

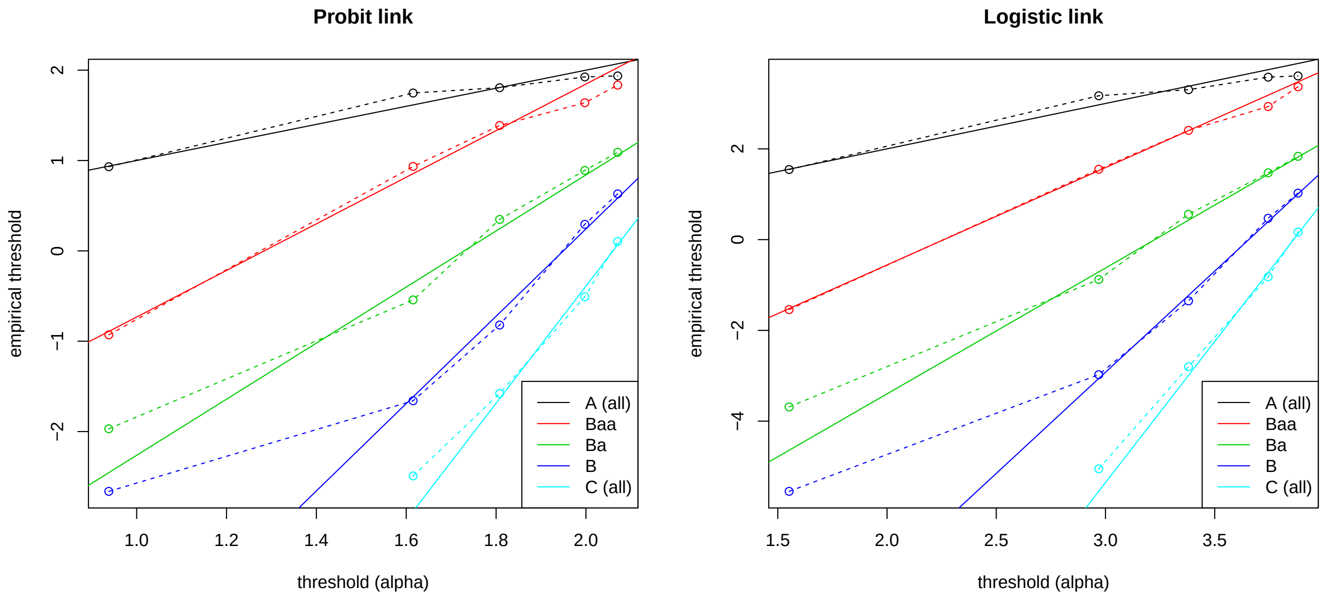

Figure 2: Goodness-of-fit of the cumulative link model to 5-year transition matrix for probit link (left panel) and logistic link (right panel). The straight lines represent the thresholds for each row (starting grade), as fitted by maximum likelihood.

than the standard normal, so the right hand panel of Figure 2 exhibits a better fit, but it seems that further adjustment is required. Albert and Chib (1991) suggest that the standard logistic is well-approximated by the t_8 distribution, and it has a coefficient of kurtosis approximately equal to that of a t_9 distribution. We have found that, for cumulative link models the fit of the logistic link (proportional odds model) compares with the fit of the t-link model for degrees of freedom between 8 and 12. Hence, the right hand panel of Figure 2 illustrates that a t-link with $\nu < 8$ may be most appropriate, an observation confirmed by the superior fit of the t-link with $\nu = 2.65$, exhibited in Figure 3.

We use maximum likelihood to find the best degrees-of-freedom for the t-link. For any degrees of freedom ν, the maximum likelihood estimate for the parameters (α, μ, σ) of (4) is obtained by maximising the log-likelihood

$$\ell(\alpha, \mu, \sigma, \nu) = \sum_{i,j} x_{ij} \log \left(F_\nu \left(\frac{\alpha_j - \mu_i}{\sigma_i} \right) - F_\nu \left(\frac{\alpha_{j-1} - \mu_i}{\sigma_i} \right) \right) \tag{6}$$

as a function of (α, μ, σ) where F_ν is the distribution function of a standard t_ν distribution. We denote the corresponding maximising values as $(\hat{\alpha}(\nu), \hat{\mu}(\nu), \hat{\sigma}(\nu))$. The profile log-likelihood function for ν,

$$\ell_p(\nu) \equiv \ell(\hat{\alpha}(\nu), \hat{\mu}(\nu), \hat{\sigma}(\nu), \nu)$$

can then be plotted, and the maximising value of ν identified. This is arguably more cumbersome than simply maximising directly over $(\alpha, \mu, \sigma, \nu)$, but we find the profile log-likelihood to be informative about the sensitivity to mis-specification of the link function. In Figure 4, we present the profile log-likelihood for the 2006-11 transition matrix analysed in Figures 2 and 3. The profile log-likelihood can be seen to be maximised at $\hat{\nu} = 2.65$, with a steep drop-off, for values of $\nu < \hat{\nu}$ and a more gradual decline for $\nu > \hat{\nu}$. A 95% confidence interval for ν can be

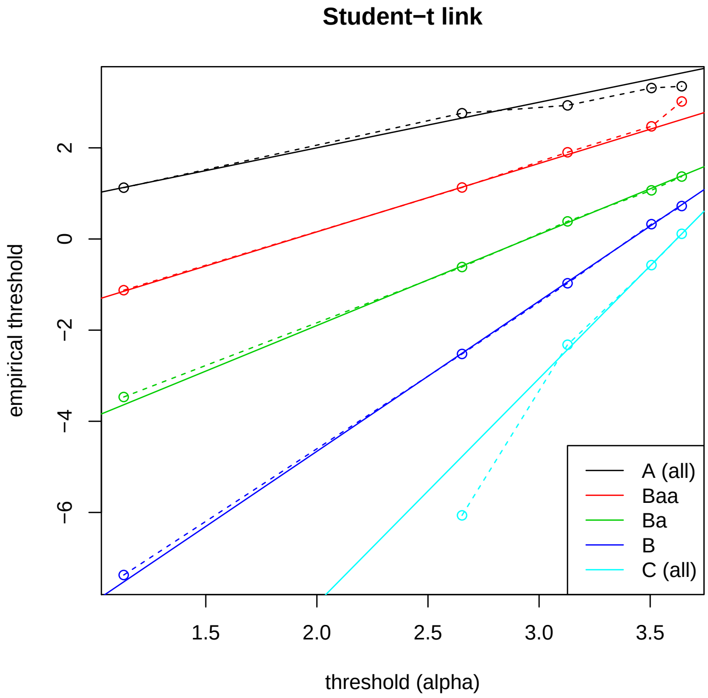

Figure 3: Goodness-of-fit of the cumulative link model to 5-year transition matrix for the t link with 2.65 degrees of freedom. The straight lines represent the thresholds for each row (starting grade), as fitted by maximum likelihood.

derived as

$$\nu \in C = \{\nu : 2\,(\ell_p(\hat{\nu}) - \ell_p(\nu)) < 3.84\}$$

which is $(2.08, 3.52)$ here. In particular, there is strong evidence in favour of a t-link with low (about 3) degrees of freedom, over the logistic or probit links which fit significantly worse.

The apparent superiority of the t-link over the logistic and, in particular, probit link raises the question of why a heavy-tailed link, corresponding to a heavy-tailed latent asset value distribution seems to be required. In Section 1, we described how a cumulative probit model was approximately equivalent to a structural model for a latent asset-value variable with independent normally distributed increments. Here, the model is fitted to aggregate data, so the correspondence is valid under the assumption that the asset value increments in the population can be modelled as marginally identically distributed normal variables. The assumption of marginal identical normal distribution applies even if individual asset value increments have different means λ_k (or equivalently that the drift parameter in the individual asset value processes is different) provided that the mean (drift) parameter λ_k is normally distributed in the population (with mean λ), However, this is not the case if the population of obligors is heterogeneous with respect to the variance of individual asset value increments. Then, the marginal distribution is a scale mixture of normals. For a discussion of normal mixture models for credit defaults, see McNeil et al (2005, Chapter 8). The most tractable normal scale-mixture model is where the individual obligor asset increment variances are assumed to have an inverse chi-squared distribution in the population. For an individual obligor, k, this implies that the latent asset-value process $\{Z_{kt}\}$ follows

$$Z_{k\,t+1}\,|\,Z_{kt} = z \sim N(z + \lambda, \tau_k^2)$$

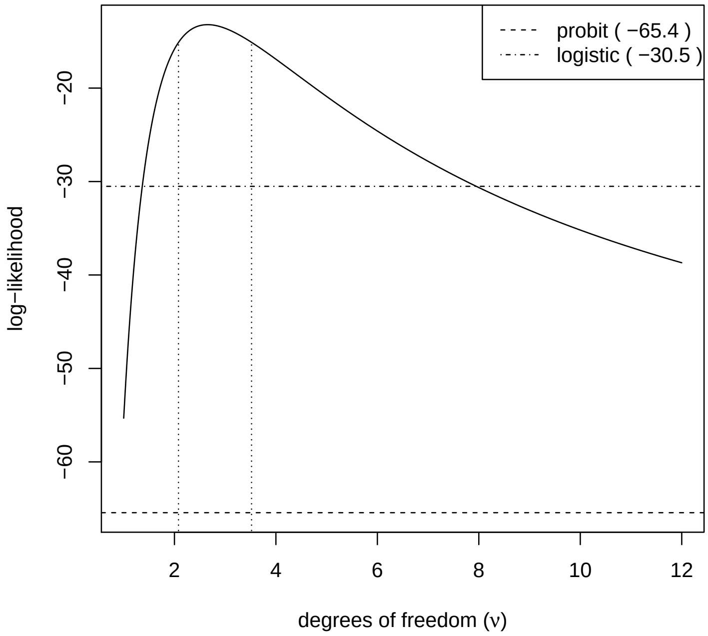

Figure 4: Profile log-likelihood for the degrees of freedom parameter ν for a cumulative t-link model. The maximum likelihood estimate is $\hat{\nu} = 2.65$, with a 95% profile likelihood confidence interval $(2.08, 3.52)$ (displayed as vertical dotted lines). The maximised likelihoods for the probit ($\nu \to \infty$) and logistic ($\nu \approx 8$) links are also displayed.

where the population distribution of the individual obligor asset increment variances, τ_k^2, is

$$\frac{\sigma^2}{\tau_k^2} \sim \chi_\nu^2.$$

Then, the (discretely observed) marginal asset value increment process is

$$\frac{Z_{k\,t+1} - (z + \lambda)}{\sigma} \mid Z_{kt} = z \sim t_\nu.$$

Hence, there is a sound theoretical basis for using a t-link model for a credit-grade transition matrix. Note that the marginal process is no longer independent in time, as the shared (unobserved) τ^2 across time for each obligor induces dependence in the individual series. In other words, although, for example, $Z_{t+1} - Z_t$ and $Z_t - Z_{t-1}$ are conditionally independent given τ^2, they are not conditionally independent given σ^2 as marginalisation over the inverse chi-squared distribution of τ^2 induces correlation in the magnitude of the increments (increments of large absolute value are associated with large values of the unobserved asset increment variance τ_k^2).

4 Further analysis

We fitted the t-link model to our credit card portfolio (monthly transitions with $D = 12$) and to our artificially created portfolio of corporate assets (annual transitions with $D = 8$). In both cases, we found that the likelihood was maximised for degrees of freedom, ν, between 1.5 and 4 across the set of transition matrices analysed. Figure 5 displays how the optimal ν varies as a function of the starting year for each transition.

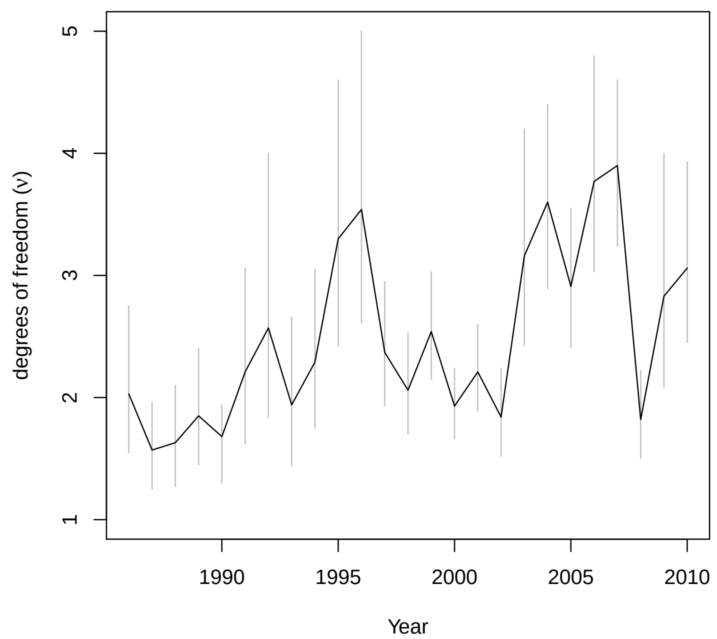

Figure 5: Maximum likelihood estimate $\hat{\nu}$ for the degrees of freedom parameter of a cumulative t-link model for a 1-year transition matrix, plotted against starting year. The vertical lines represent 95% confidence intervals, based on profile likelihood.

We also investigated how the variability in $\hat{\nu}$ compares with the variability in marginal default rate in the portfolio across time. Figure 6 displays the time series of $1/\hat{\nu}$, superimposed over the time series of the (empirical) probability of flow to default. It is interesting to note that $1/\hat{\nu}$ is higher, representing a greater portfolio heterogeneity in years where the default rate is higher, and correspondingly troughs in default rate correspond to the periods of greatest portfolio homogeneity.

Goodness-of-fit of the cumulative t-link model can be evaluated by comparing it directly against the saturated (unstructured) model which simply models the rows of each transition matrix as a set of unrelated unconstrained discrete probability distributions over the sample space $\{1, \ldots, D\}$. Model comparison is based on the log-likelihood ratio statistic

$$L = 2\left(\ell(\hat{\alpha}, \hat{\mu}, \hat{\sigma}, \hat{\nu}) - \ell_S(\hat{p})\right)$$

where ℓ is given by (6),

$$\ell_S(p) = \sum_{i,j} x_{ij} \log p_{ij}$$

and

$$\hat{p}_{ij} = \frac{x_{ij}}{\sum_{k=1}^{D} x_{ik}}.$$

A formal log-likelihood ratio comparison of the cumulative link and saturated models, rejects the former in favour of the latter when L is too large as calibrated under the null distribution χ^2_d,

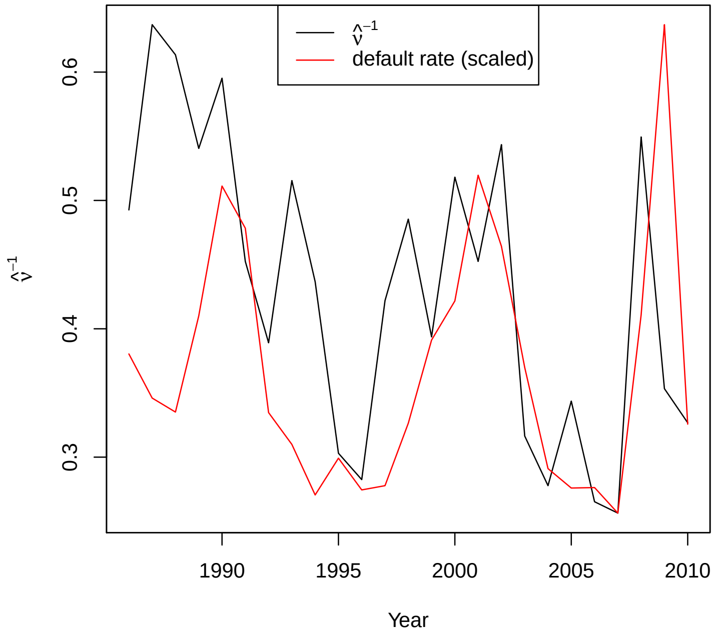

Figure 6: Maximum likelihood estimate $1/\hat{\nu}$ for the inverse degrees of freedom parameter of a cumulative t-link model for a 1-year transition matrix, plotted against starting year, together with the marginal default rate, also plotted against starting year (y-axis suppressed).

where the degrees of freedom is the difference in the dimensionality of the parameter spaces. For the scale-varying cumulative t-link model (4) for a $D \times D$ transition matrix with an absorbing state (D) there are $(D-1)$ ordinal probability distributions to estimate, corresponding to the $(D-1)$ possible non-default initial states. A saturated model for each of these D-category ordinal probability distributions involves estimating $(D-1)$ free probabilities. Hence the dimensionality of the saturated model is $(D-1)^2$. For our cumulative link model, the free parameters are the $D-1$ thresholds, $\{\alpha_j, 1 \leq j < D\}$, the row-specific mean and variance parameters $\{(\mu_i, \sigma_i^2), i > 1\}$ and the degrees of freedom ν. Hence the degrees of freedom for our cumulative t-link model is

$$\begin{aligned} d &= (D-1)^2 - ((D-1) + 2(D-2) + 1) \\ &= (D-2)(D-3) - 1. \end{aligned}$$

Alternatively, one can compare the models using the AIC or BIC which prefer the cumulative link model to the saturated model when $L < 2d$ and $L < d \log n$ respectively, where n is the sample size. These approaches are arguably preferable for transition matrix analysis where the matrices under analysis may be sparse (as is the case here).

Figure 7 displays the log-likelihood ratio statistic for each of the 25 1-year transition matrices, together with the thresholds for comparison between the cumulative link model and saturated model using both the likelihood ratio test and AIC. Where the threshold is exceeded the saturated model is preferred. The threshold for BIC is not displayed, as it exceeds 200 in every year. Hence, BIC strongly favours the cumulative link model, AIC almost always favours the cumulative link model (except 2000, 2002), and the likelihood ratio test results are more mixed, with the saturated model preferred in 9 of the 25 years. Given the reservations about the likelihood ratio test in this analysis, described above, we are happy to conclude that the scale-varying cumulative t-link model fits these transition matrices well, in comparison with the unstructured alternative. Although we do not present the results here, the scale-varying cumulative t-link model is also preferred over simpler alternatives, such as the constant-scale model (3) or models with logistic or probit link functions.

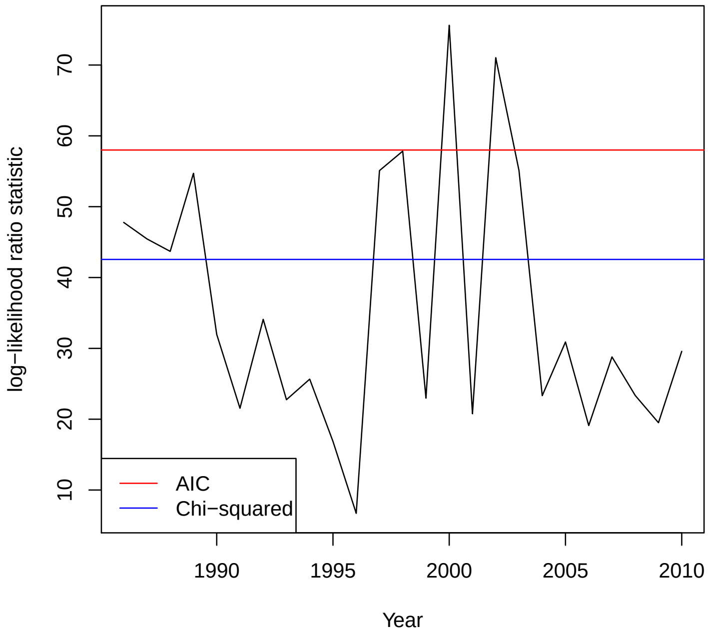

Figure 7: Maximised log-likelihood for a cumulative t-link model for a 1-year transition matrix, plotted against starting year. The thresholds for comparison against the saturated (unstructured) model are also displayed when comparison is based on AIC and likelihood ratio chi-squared.

5 Conclusions and further work

When modelling a portfolio credit grade transition process, natural heterogeneity between the obligors comprising the portfolio makes a heavy-tailed link such as Student-t, attractive when considering a cumulative link model. Our experience, with both corporate and retail portfolio data, suggests that degrees of freedom ν between 2 and 3 may be appropriate, but that the conservative approach is to choose something higher in this range, as the consequences in terms of goodness-of-fit are more serious for underspecification of ν than (mild) overspecification. However, Figure 6 suggests there may be scope for using a varying degrees of freedom, modelled in conjunction with other cumulative link model parameters. One possibility would be to consider a joint model where there is a direct relationship between ν and other model parameters, or where ν and other model parameters co-vary as a function of explanatory variables, such as relevant macroeconomic covariates.

For a complete forecasting model, it is necessary to augment this model for individual historical transitions so that future transition dynamics can be predicted. This typically requires further modelling assumptions. For example, one might make the further assumption that the credit grade migration process is Markov, which implies that, for any $t < u < v$

$$P(t, v) = P(t, u)P(u, v).$$

Hence, under the Markov assumption, it is only necessary to forecast transition matrices between consecutive time points. [With the further assumption, of time-homogeneity, under which $P(t, u)$ is a function of $(t - u)$, then only a single transition matrix forecast is necessary, but such an assumption ignores the natural response of the grade migration process to the underlying economic cycle]. The Markov model may also be sensitive to portfolio heterogeneity, in that a plausible Markov model at individual obligor level, with different underlying parameters may not be well approximated by assuming the population-averaged parameters obtained from a portfolio-level analysis. Malik and Thomas (2012) for retail credit grade migration and Lando and Skødeberg (2002) for corporate data find significant evidence of non-Markov behaviour in credit grade transition. Incorporating the t-link into a non-Markov model is the subject of ongoing research.

Finally, we showed in Section 4 that the cumulative t-link model fits the corporate portfolio data well. For some retail portfolios we have examined, the t-link model, although vastly superior to links such as logistic and probit with lighter tails, does not always compare well with the saturated alternative, even for a comparison using BIC. The issue seems to be that there is a natural skewness in the data with transitions to better quality credit grades exhibiting longer tails than transitions to inferior grades. Preliminary investigation has shown that this can be effectively modelled using a skew t-distribution. Various families of skew-t are available. We have found that a cumulative link model where the link is based on the distribution function of the skew-t distribution of Azzalini and Capitanio (1985) provides a superior fit to the data. On the other hand, Aas and Hobæk Haff suggest a skew t which is a member of the generalised hyperbolic family (see also McNeil et al, 2005, Section 3.2.3). The characterisation of the generalised hyperbolic as a mixture of normals suggests that a cumulative link model for a portfolio transition matrix, based on this kind of skew t latent asset value variable, can be motivated as an inhomogeneous mix of Merton models for individual obligors. This is also an area of ongoing research.

References

Aas, K. and Hobæk Haff , D. (2006). The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, **4**, 275-309.

Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669-679.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society B*, **65**, 367389.

Chan, N. H., Wong, H. Y. and Zhao, J. (2012). Structural model of credit migration. *Computational Statistics and Data Analysis*, **56**, 3477-3490.

Christensen, R.H.B. (2012) Analysis of ordinal data with cumulative link models – estimation with the R-package `ordinal`.
Available at `http://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf`

Feng, D., Gourieroux, C. and Jasiak, J. (2008). The ordered qualitative model for credit rating transitions. *Journal of Empirical Finance*, **15**, 111-130.

Hu, Y-T., Kiesel, R. and Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings. *Journal of Banking and Finance*, **26**, 1383-1406.

Kizilkaya,K., Carnier, P., Albera, A., Bittante, G. and Tempelman, R.J. (2003). Cumulative t-link threshold models for the genetic analysis of calving ease scores. *Genetics Selection Evolution*, **35**, 489-512.

Lando, D. and Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*, **26**, 423-444.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society B*, **42**, 109-142.

McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management.* Princeton University Press.

McNeil, A.J. and Wendin, J.P. (2006). Dependent credit migrations. *Journal of Credit Risk*, **2**, 87-114.

Malik, M. and Thomas, L.C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, **28**, 261-272.

Merton R.C. (1974) On the pricing of corporate debts: the risk structure of interest rates. *Journal of Finance*, **29**, 449-470.

Nickel, P., Perraudin, W. and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking and Finance*, **24**, 203-227.

Ntzoufras, I., Dellaportas, P. and Forster, J.J. (2003). Bayesian variable and link determination for generalised linear models. *Journal of Statistical Planning and Inference*, **111**, 165-180.