Estimating Default Correlations from Short Panels of Credit Rating Performance Data[1]

Michael Gordy Erik Heitfield

January 29, 2002

[1]The opinions expressed here are those of the authors, and do not reflect the views of the Board of Governors or its staff. Michael Gordy can be reached by phone at 202-452-3705 and by e-mail at Michael.Gordy@frb.gov. Erik Heitfield can be reached by phone at 202-452-2613 and by e-mail at Erik.Heitfield@frb.gov. Written correspondence to either author should be sent to Mail Stop 153, Federal Reserve Board, Washington, DC 20551 USA.

Abstract

When models of portfolio credit risk are calibrated to historical ratings performance data, parameters that capture cross-obligor dependence can be (and often are) fit directly to estimated default correlations. The accuracy of our measures of credit value-at-risk therefore rests on the precision with which default correlations can be estimated. In practice, data are always scarce. The rating system may cover many obligors, but performance data span, at most, two or three decades. Nonetheless, the moments-based estimators commonly used by practitioners make minimal use of parameter restrictions. In this paper, we demonstrate that these estimators perform quite badly on the sample sizes typically available, and generally produce a large downward bias in estimated default correlations. Models calibrated in this manner are thus likely to understate value-at-risk quite significantly.

Our main theme concerns the trade-off between precision and robustness in the calibration of default correlation parameters. We show how economically meaningful assumptions about the unobserved process that determines when obligors default can be used to generate natural restrictions on cross-obligor default correlations. We demonstrate how these restrictions can be imposed when maximum likelihood methods are used to calibrate parameters to historical ratings performance data, and then assess the associated improvements in small sample behavior of the estimators. We apply our various estimators to Moody's and S&P ratings performance data.

1 Introduction

In well diversified loan portfolios, aggregate credit risk is driven by correlations in defaults and ratings changes across obligors. Holding the composition of loans in a portfolio constant, if credit events are largely independent then portfolio losses can be easily and precisely forecast. As cross-obligor correlations increase, our ability to diversify away credit risk at the aggregate level is reduced, so the greater the year-to-year variability in aggregate performance the greater the value-at-risk. When models of portfolio credit risk are calibrated to historical ratings performance data, parameters that capture cross-obligor dependence can be (and often are) fit directly to estimated default correlations. The accuracy of our measures of credit value-at-risk therefore rests on the precision with which default correlations can be extracted.

In measuring cross-obligor default correlations data scarcity is a serious and unavoidable problem. Though publicly and privately available databases can provide information on the default histories of large numbers of rated obligors, these data span, at most, just a few decades. Without long time-series of performance data there may be significant uncertainty in measurements of the default probabilities associated with individual rating grades. These uncertainties are greatly magnified for estimates of default correlations. Parametric restrictions serve as a partial substitute for data, but offer no free lunch. If the restrictions arise naturally from the underlying economics of credit risk, then they allow more precise estimation of model parameters from a given amount of data. However, the more powerful the restrictions, the greater the risk that the model will oversimplify reality, and thereby produce less robust estimators.

This paper shows how economically meaningful assumptions about the unobserved process that determines when obligors default can be used to generate restrictions on cross-obligor default correlations. We demonstrate how these restrictions can be imposed by using maximum likelihood methods to estimate model parameters from historical ratings performance data. We explore the small sample properties of these estimators, and apply them to data on Moody's and S&P ratings.

Starting from a simple but widely used structural default model, Section 2 shows how parametric restrictions on default correlations naturally follow from assumptions about the economic factors driving changes in obligor asset values. We assume defaults are well described by a simple two-state (default/no default) Merton (1974) default model similar to that underlying the popular CreditMetrics credit risk model (Gupton, Finger and Bhatia 1997). In this framework, each obligor earns some random return on its assets at the horizon. If the return is sufficiently negative, the value of the obligor's assets falls beneath the value of its fixed liabilities, and the obligor defaults. The realized returns are a weighted sum of a set of common risk factors (representing systematic risk) and a shock idiosyncratic to an obligor. Default correlations are determined by obligors' "factor loadings," which measure sensitivity of asset values to the common risk factors. Limiting the number of common risk factors and restricting the way factor loadings vary among obligors generates economically meaningful restrictions on cross-obligor default correlations.

Section 3 shows how factor loadings and other parameters of the structural model can be estimated from ratings performance data using maximum likelihood methods. Because they depend on nonlinear optimization techniques the maximum likelihood estimators we propose are somewhat more computationally demanding than more commonly used method of moments estimators, but unlike method of moments estimators they easily lend themselves to imposing structural parameter restrictions.

Assuming the structural default model is correctly specified the estimators we propose will be asymptotically consistent. Loosely speaking this means that estimated parameters will get closer and closer to the true parameter values as the number of years of rating performance data gets increasingly large. Unfortunately, in practice we are unlikely to observe ratings performance data over more than thirty years. Given this data limitation, there is no guarantee maximum likelihood will produce unbiased parameter estimates. Section 4 presents results for a Monte Carlo study of the small sample properties of three different maximum likelihood estimators as well as a simple method of moments ("MM") estimator.

The three maximum likelihood estimators we study can be ordered by the restrictiveness of the assumptions they impose. The first, and most general, MLE allows for the possibility that obligors in different rating grades may be sensitive to different risk factors. The second MLE imposes the restriction that obligors in all rating grades are sensitive to a single unique systematic risk factor, but allows factor loadings to vary across grades. Finally, the most restrictive MLE requires that factor loadings be constant across rating grades.

If the restrictions imposed by the last estimator are correct, the three MLE estimators and MM estimator all are consistent. However, our Monte Carlo simulations show that all the estimators generate downward-biased factor loadings when small samples are used. Not surprisingly, this bias is most severe for the least restrictive estimators. We find that the biases associated with estimators that assume a single systematic risk factor, while measurable, are much smaller than those associated with estimators that allow for the possibility of multiple systematic risk factors.

The findings from our Monte Carlo study suggest that by imposing a single systematic risk factor assumption but allowing factor loadings to vary smoothly across different types of obligors, one can obtain maximum likelihood estimators that are not significantly biased in small samples but nonetheless permit a reasonable degree of flexibility in the structure of default correlations across obligors. In Section 5 we apply several such estimators to Moody's and S&P ratings. Because the aggregate ratings data we use only provides information on the credit rating and default experience of obligors, we assume that an obligor's factor loading depends only on its rating.

Arguments can be made for assuming any number of different functional relationships between credit quality measures and factor loadings. For example, it seems reasonable to expect that large obligors with more diversified business lines are more sensitive to macroeconomic fluctuations than small obligors whose risks are more idiosyncratic. If large obligors are also higher quality, then we should observe that factor loadings are a decreasing function of default probabilities. On the other hand, we might view lower quality obligors as simply higher quality obligors who have fallen on hard times. In the context of a one risk factor model, we cannot reject the hypothesis that factor loadings are, in fact, constant with respect to credit quality.

2 A Structural Default Model

In this analysis we work within a two-state version of the widely-used CreditMetrics framework. Assume we have a set of obligors, indexed by i. Associated with each obligor is a latent variable Y_i which represents the normalized return on an obligor's assets. Y_i is given by

$$Y_i = \mathbf{Z}\omega_i + \xi_i \epsilon_i. \tag{1}$$

where $\mathbf{Z}$ is a K-vector of systematic risk factors. These factors capture unanticipated changes in economy-wide variables such as interest rates and commodity prices that affect asset returns across all industries. We assume that $\mathbf{Z}$ is a mean-zero normal random vector with variance matrix Ω. We measure the sensitivity of obligor i to $\mathbf{Z}$ by a vector of factor loadings, ω_i. ϵ_i represents obligor-specific risk. Each ϵ_i is assumed to have a standard normal distribution and is independent across obligors and independent of $\mathbf{Z}$. Without loss of generality, the covariance matrix Ω is assumed to have ones on the main diagonal (so each Z_k has a standard normal marginal distribution), and the weights ω_i and ξ_i are scaled so that Y_i has a mean of zero and a variance of one. The obligor defaults if Y_i falls below the default threshold γ_i. By construction, then, the unconditional probability of default ("PD") of obligor i is equal to the standard normal CDF evaluated at γ_i.

To allow the model to be calibrated using historical data of the sort available from the rating agencies, we group the obligors into G homogeneous "buckets" indexed by g. In the applications that follow the buckets comprise an ordered set of rating grades. In principle, however, a bucketing system can be defined along multiple dimensions. For example, a bucket might be composed of obligors of a given rating in a particular industry and country. Within a bucket, each obligor has the same default threshold γ_g so that the PD of any obligor in grade g is

$$\bar{p}_g = \Phi(\gamma_g), \tag{2}$$

where $\Phi(z)$ is the standard normal CDF.

The vector of factor loadings is assumed to be constant across all obligors in a grade. so we can re-write the equation for Y_i as

$$Y_i = X_g w_g + \epsilon_i \sqrt{1 - w_g^2}. \tag{3}$$

where

$$X_g = \frac{\sum_k Z_k \omega_{g,k}}{\sqrt{\omega_g' \Omega \omega_g}}$$

is a univariate bucket-specific common risk factor. By construction, each X_g is normally distributed with mean zero and unit variance. The G-vector $\mathbf{X} = (X_1, \ldots, X_G)$ has a multivariate normal distribution. Let σ_{gh} denote the covariance between X_g

and X_h. In general, we expect $\sigma_{gh} > 0$. The factor loading on X_g for obligors in bucket g is

$$w_g = \sqrt{\omega_g' \Omega \omega_g},$$

which is bounded between zero and one. We eliminate ξ_i from equation (1) by imposing the scaling convention that the variance of Y_i is one.

The advantage of writing Y_i in terms of X_g and w_g rather than $\mathbf{Z}$ and ω_g is that we then only need to keep track of one risk factor per bucket. We can think of X_g as summarizing the total effect of $\mathbf{Z}$ on obligors in bucket g, and w_g as describing the sensitivity of those obligors to the bucket-specific common risk factor. In the discussion that follows, the term *risk factors* should be taken to refer to X_g. The term *structural risk factors* will be used to identify the elements of $\mathbf{Z}$ because they reflect underlying economic variables. Likewise *factor loadings* will refer to w_g and *structural factor loadings* will refer to ω_g.

Correlation in default rates across obligors is driven by correlations in the risk factors. It can be shown that the correlation in defaults between an obligor in bucket g and an obligor in bucket h is

$$\rho_{gh} = \frac{F(\gamma_g, \gamma_h; w_g w_h \sigma_{gh}) - \bar{p}_g \bar{p}_h}{\sqrt{\bar{p}_g(1-\bar{p}_g)}\sqrt{\bar{p}_h(1-\bar{p}_h)}} \quad (4)$$

where $F(z_1, z_2; s_{12})$ denotes the joint CDF for a mean-zero bivariate normal random vector with unit variances and covariance s_{12}. In the special case where both obligors lie in the same bucket, the within-bucket default correlation is

$$\rho_g = \frac{F(\gamma_g, \gamma_g; w_g^2) - \bar{p}_g^2}{\bar{p}_g(1-\bar{p}_g)}. \quad (5)$$

Given sufficient data, it is possible to estimate all $G(G+1)/2$ default correlations defined by equations (4) and (5). However, when data are scarce many of these parameters may be unidentified or poorly identified, obviating the need to limit the number of parameters that are estimated. This can be accomplished by imposing *ex ante* restrictions on the factor loadings and risk factor correlations.

Although our default model assumes within-bucket homogeneity and imposes explicit distributional assumptions on the common and idiosyncratic risk factors, in some respects it is actually quite general. Importantly, no restrictions are imposed on the correlation in default rates across buckets. In this regard the model is, in fact, more general than the credit risk model underlying the proposed Basel II internal ratings based capital standard. IRB capital charges are *portfolio-invariant*. That is, the charge for a lending instrument depends only on its own properties, and not those of the portfolio in which it is held. Gordy (2001) shows that portfolio-invariance requires that there be a *single* common systematic risk factor X for all buckets.

Assuming that correlations in asset values are driven by a single systematic risk factor is equivalent to imposing the following restriction in the parameterization of ρ_{gh}.

R 1 (One Risk Factor) *$\sigma_{gh} = 1$ for all (g, h) bucket pairs.*

R1 is equivalent to requiring that $X_1 = X_2 = \ldots = X_G$. A sufficient condition for R1 is that there is exactly one structural risk factor (i.e. $K = 1$). One can easily imagine circumstances under which this condition will not hold. For example asset values for obligors in different industries or countries likely depend on different structural risk factors. Nonetheless, if a portfolio is relatively homogeneous, or if sectoral distinctions among obligors cannot be observed from available data, this restriction can serve as a reasonable approximation.

While R1 imposes a restriction on the correlation among reduced form risk factors, it does nothing to restrict the sensitivity each obligor's asset value to those factors. A different reduced form factor loading is associated with each bucket, and no restrictions are imposed on the cross bucket relationship between factor loadings. In practice it may be reasonable to assume that factor loadings vary smoothly with obligor default probabilities (or equivalently with obligor default thresholds). This assumption can be imposed by expressing factor loadings as a continuous function of default thresholds.

R 2 (Smooth Factor Loadings) *$w_g = \Lambda(\lambda(\gamma_g))$ for all g, where $\Lambda(\cdot)$ is a continuous, strictly monotonic link function that maps real numbers onto the interval [-1,1] and $\lambda(\cdot)$ is a continuous index function that maps default thresholds onto the real line.*

The choice of the link function is rather arbitrary. In the analysis that follows we use the simple arctangent transformation

$$\Lambda(\lambda) = \frac{2}{\pi} \arctan(\lambda).$$

This function is linear with unit slope in a neighborhood of $\lambda = 0$ and asymptotes smoothly toward positive (negative) one as λ approaches positive (negative) infinity. The specification of the index function is more important than the choice of the link because it can be used to restrict the way w varies with γ. If the index function is monotonic in γ than mapping from γ to w will be monotonic as well. The more parsimonious is the index function, the more restrictive is the implied relationship between the default thresholds and the factor loadings.

The strongest restriction one can impose on the factor loadings is to assume that they are constant across all obligors.

R 3 (Constant Factor Loading) $w_g = w_h$ *for all* (g, h) *bucket pairs.*

Together, R1 and R3 imply that the structural factor loadings are constant across buckets. Note that R3 is a special case of R2 in which the index function $\lambda(g)$ is a constant.

3 Maximum Likelihood Estimators

For the remainder of this paper, we assume that we have access to historical performance data for a credit ratings system. For each of T years and G rating grades, we observe the number of obligors in grade g at the beginning of year t (a "grade-cohort"), and the number of members of the grade-cohort who default by year-end. We assume that the default threshold γ_g and the factor loading w_g are constant across time for each rating grade, and that the vector of risk factors $\mathbf{X}$ is serially independent. The task at hand is to estimate γ_g and w_g for each rating grade. Given these parameter estimates we can recover PDs and default correlations using equation (2), (4), and (5).

Let n_g and d_g denote the number of obligors and the number of defaults in grade g, and let $\mathbf{n}$ and $\mathbf{d}$ be the corresponding G-vectors containing data for all grades. Throughout this derivation, we assume the random process that generates the observed number of obligors in each grade is independent of the process that generates the observed number of defaults in each grade. This allows us to treat $\mathbf{n}$ as a fixed parameter in the likelihood function for $\mathbf{d}$.

Conditional on X_g defaults in grade g are independent, and each default event can be viewed as the outcome of a Bernoulli trial with success probability

$$p_g(X_g) = \Phi\left(\frac{\gamma_g - w_g X_g}{\sqrt{1 - w_g^2}}\right). \tag{6}$$

The total number of defaults conditional on X_g follows the binomial distribution

$$L(d_g|X_g) = \binom{n_g}{d_g} p_g(X_g)^{d_g} \left(1 - p_g(X_g)\right)^{n_g - d_g}. \tag{7}$$

Since defaults are conditionally independent across grades, the joint likelihood of $\mathbf{d}$ conditional on $\mathbf{X}$ is simply the product of the G conditional likelihoods defined in (7). The unconditional likelihood for $\mathbf{d}$ is thus,

$$L(\mathbf{d}) = \int\limits_{\Re^G} \prod_{g=1}^{G} \left(\binom{n_g}{d_g} p_g(x_g)^{d_g} \left(1 - p_g(x_g)\right)^{n_g - d_g}\right) dF(\mathbf{x}). \tag{8}$$

$F(\mathbf{x})$ is the multivariate normal CDF of $\mathbf{X}$. Equation (8) is a function of the parameters $\mathbf{w} = (w_1, \ldots, w_G)$, $\gamma = (\gamma_1, \ldots, \gamma_G)$, and Σ, the variance matrix of $\mathbf{X}$ containing $(G-1)G/2$ free covariance parameters.

In principle, we could maximize the product of (8) across T observations with respect to all $2G + (G-1)G/2$ free parameters simultaneously. This would provide unrestricted full information maximum likelihood estimates of the parameters. In practice, however, this strategy is computationally feasible only when G is small. Unless the common factor covariance parameters are of particular interest, a limited information approach that does not involve estimating the elements of Σ is preferable. Integrating X_g out of equation (7) yields the marginal likelihood

$$L(d_g) = \int\limits_{\Re} \binom{n_g}{d_g} p_g(x)^{d_g} \left(1 - p_g(x)\right)^{n_g - d_g} d\Phi(x). \tag{9}$$

This function depends only on the parameters w_g and γ_g, so estimates of $\mathbf{w}$ and γ can be obtained by maximizing the marginal likelihood for each grade, one grade at a time. This procedure yields our least restrictive maximum likelihood estimator that imposes no restrictions in the parameters of the default model described in Section 2. Because this estimator does not utilize information about the potential correlation in default rates across grades, it is not asymptotically efficient (except in the unrealistic special case where $\sigma_{gh} = 0$ for all $g \neq h$).

R1 implies that the effect of $\mathbf{X}$ on all obligors can be represented by a single standard normal *scalar* variable X. Under this restriction we can rewrite (8) as

$$L(\mathbf{d}) = \int_{\Re} \prod_{g=1}^{G} \left(\binom{n_g}{d_g} p_g(x)^{d_g} \left(1 - p_g(x)\right)^{n_g - d_g} \right) d\Phi(x). \tag{10}$$

Maximizing this likelihood over $\mathbf{w}$ and γ yields a full information likelihood estimator that imposes the one risk factor restriction.

Rather than estimate the elements of $\mathbf{w}$ directly one can substitute the formula in R2 into equation (10) and maximize the resulting equation over γ and the parameters of the index function $\lambda(\gamma)$. This procedure yields a FIML estimator that imposes both the one risk factor and the smooth factor loading restrictions. Similarly, R1 and R3 can be imposed by replacing the vector $\mathbf{w}$ in equation (10) with a single loading w and maximizing the resulting likelihood with respect to γ and the scalar w.

If both R1 and R3 hold, then all the maximum likelihood estimators described in this section are consistent. Furthermore, the estimator that imposes R1 and R3 is efficient in the sense that it achieves the lowest possible asymptotic variance among consistent estimators. It is important to emphasize, however, that in finite samples some or all of these maximum likelihood estimators may be biased. In the next section we use Monte Carlo simulations to investigate the small sample properties of these estimators and to compare them with a simple method of moments estimator used in previous research.

4 Monte Carlo Simulations

If many decades of ratings performance data were available, the asymptotic results of the previous section would pose a clear trade-off. On one hand, the more restrictive maximum likelihood estimators yield more precise estimates if the restrictions they impose are valid; on the other hand, the less restrictive estimators are more robust to specification errors. When ratings performance data are in short supply (i.e. T is small) the tradeoff becomes more complicated because less efficient, more robust estimators may also be biased. Monte Carlo simulations provide a method for studying the potential small sample biases associated with the estimators we propose.

The following four estimators are examined in this analysis.

- MM – an unrestricted method of moments estimator.
- MLE1 – the limited information maximum likelihood estimator.
- MLE2 – the full information maximum likelihood estimator that imposes R1.
- MLE3 – the full information maximum likelihood estimator that imposes R1 and R3.

MM has been used in previous research and is described in detail by Gordy (2000). It is included in this analysis to provide a benchmark for assessing the maximum likelihood estimators.

In each Monte Carlo simulation, we constructed a synthetic dataset intended to represent the type of historical data available from the major rating agencies. Data were simulated for three rating grades. Grade "A" corresponds to medium to low investment grade (S&P A/BBB), grade "B" corresponds to high speculative grade (S&P BB), and grade "C" corresponds to medium speculative grade (S&P B). Table 1 summarizes characteristics of these three grades.[1] Simulated defaults in each grade were generated according to the stochastic model described in Section 2 with R1 and R3 imposed.

Two sets of Monte Carlo simulations were undertaken. In the first, 500 synthetic datasets were generated for four different values of T: 20, 40, 80, and 160. In each case a "true" factor loading of 0.45 was assumed. These simulations were intended to shed light on the properties of our estimators as the number of years of default data increase. Though estimates of both factor loadings and default thresholds were obtained for each simulated dataset, we will postpone discussing default thresholds for the time being. Table 2 summarizes the means, standard deviations, and root mean squared errors ("RMSE") for the estimates of w given each of the four sample sizes.[2] Figure 1 displays the median and the 5th and 95th percentiles of the estimated parameter values.

[1]S&P grade-cohorts are somewhat larger than we have assumed, but are similar in the relative preponderance of higher grade obligors.

[2]Root mean squared error measures the geometric distance between estimated parameter values and the true values and takes into account both average bias and variance. It is defined as follows. Let $\hat{w}_i$ denote the parameter estimate generated from the ith Monte Carlo simulation, let w_0 denote the true parameter and let N be the total number of simulations. RMSE $= \left(\frac{1}{N}\sum_i(\hat{w}_i - w_0)^2\right)^{1/2}$.

Not surprisingly, properties of all four estimators improve as T increases. The means become closer to 0.45 and the variances and RMSEs decrease. Also as expected, for large values of T the more restrictive estimators are more tightly clustered around 0.45 than the less restrictive estimators. These findings simply verify the asymptotic results described in Section 3. More surprising is the rather poor performance of MM and MLE1 when T is small. Though all four estimators appear to be downward-biased in small samples, the bias of MM and MLE1 is substantially worse than that of MLE2 and MLE3.

In real-world applications, we could never hope to observe 80 or 160 years of default data. In fact, for most applications $T = 20$ is probably optimistic. S&P historical performance data, described in Brand and Bahar (1999), contain less than 20 years of data. Moody's performance data go back to 1970, but there is believed to be an important break in the time-series at 1983 due to a change in Moody's rating methods. Banks' internal rating systems typically contain even shorter time-series, though larger grade-cohorts. To explore the small-sample properties of our estimators in greater detail, a second set of Monte Carlo simulations was run with T fixed at 20. Four groups of 1,500 synthetic datasets were simulated for a grid of "true" factor loadings from 0.15 to 0.60.[3]

Table 3(a) and 3(b) show the distributions of estimated default thresholds and implied default probabilities. Even when T is small, all four estimators generally produce very accurate estimates of default thresholds and, therefore, of the corresponding PDs.[4]

Tables 4(a) through 4(d) describe the distributions of estimated factor loadings. Several strong patterns can be seen in these tables. Most striking, is the large downward bias associated with MM and MLE1. This problem is particularly significant for high quality grades when the true factor loadings are high. MLE2 and MLE3 are also biased downward, but the magnitude of the bias is less severe. In contrast to the

[3]For a small number of trials, the simulated data did not permit identification of all model parameters. In other trials, the optimization routines used to calculate the maximum likelihood estimators failed to converge. For these reasons, the number of estimated parameters for each configuration of parameter values was sometimes less than 1,500. See Appendix A for details on identification and convergence problems in this Monte Carlo study.

[4]As a technical aside, we note that it should be preferable to estimate default thresholds, rather than the PDs directly. Because the distribution of $\hat{\gamma}$ is more likely to be approximately symmetric, test statistics (including confidence intervals) should be well-behaved. PDs, by contrast, are bounded at zero, so estimated PDs for the higher quality grades are likely to have highly asymmetric distributions.

results for MM and MLE1, the magnitude of the bias for MLE2 does not appear to depend on the grade in any systematic way.

Based on the root mean squared error criterion, MLE3 clearly outperforms the other three estimators; and more generally, the more restrictive estimators outperform the less restrictive estimators. The greatest gain in efficiency appears to occur when R1 is imposed. Because it incorporates information on cross-grade default correlations, MLE2 produces substantially more accurate estimates of high-grade factor loadings than MLE1 or MM.

Overall, the Monte Carlo results indicate that when the number of years of default data is relatively modest, we may pay a significant price for failing to impose true parameter restrictions. The substantial small sample bias and high root mean squared error associated with unrestricted estimators must therefore be weighed against the potential bias that would result from imposing invalid parameter restrictions.

5 Estimation Using Public Ratings Data

Our Monte Carlo study demonstrates that when data are limited, failing to impose reasonable restrictions on cross-obligor default correlations can lead to significant biases in parameter estimates. Much of this bias can be eliminated by imposing a single systematic risk factor assumptions and by assuming that asset correlations vary smoothly with obligor default thresholds. In this section we apply the maximum likelihood estimators described in Section 3 to rating performance data from Moody's and S&P.

For each dataset, we estimate four models. The first imposes no restrictions and corresponds to MLE1 in the Monte Carlo study. Given our Monte Carlo results, we can reasonable expect parameter estimates from this model to have substantial biases. The second model imposes R1 and relatively flexible form or R2 in which factor loadings depend on the quadratic index function of default thresholds

$$\lambda(\gamma) = \beta_0 + \beta_1 \gamma + \beta_2 \gamma^2.$$

The third model is identical to the second, but imposes the additional restriction that the index function is linear in γ (i.e. $\beta_2 = 0$). The fourth and most restrictive model assumes a constant factor loading (i.e. $\beta_1 = \beta_2 = 0$). This model corresponds to

MLE3 in the Monte Carlo study. Estimators generated by this model should exhibit very little bias if the assumptions imposed by the model are reasonably accurate.

Data from Standard and Poor's covers the 17 year period from 1981 to 1997, while a larger dataset from Moody's covers the 29 year period from 1970 to 1998. Each dataset reports the number of obligors in a particular rating grade at the beginning of a year and the number of defaults that occur by year's end. The S&P data includes information for seven major grades from AAA to CCC, and the Moody's data includes information the seven comparable grades from Aaa to Caa. In both cases, no defaults are observed for the top two grades. Therefore the analysis of S&P data is limited to grades A through CCC, and the analysis of Moody's data is limited to grades A through Caa.

Tables 5 and 6 show estimated default thresholds for each of the model specifications described in Section 2. Standard errors are reported in parentheses. The most important fact to take away from these tables is that all four specifications generate comparable parameter estimates with similar standard errors. This is consistent with our Monte Carlo results showing that estimated default thresholds are relatively insensitive to the parameterization of factor loadings.

Estimated factor loadings are shown in Tables 5(a) and 5(b). Note that while standard errors fall as increasingly strong restrictions are imposed on the cross-grade relationship between among loadings, parameter estimates are biased if these restrictions are not valid. We can investigate whether the cross-grade parameter restrictions are consistent with observed data by comparing the restricted and unrestricted model specifications. Tables 6(a) and 6(b) report the estimated coefficients of the index function for each of the three restricted models. In both datasets, the higher order parameters in the quadratic and linear index models are not significantly different from zero.

The bottom rows of Tables 6(a) and 6(b) show the results of likelihood-ratio specification tests comparing each of the three index models with the unrestricted factor loading models. In each case the likelihood-ratio statistic tests the maintained hypothesis that the restricted model is true against the alternative hypothesis that some less restrictive relationship between factor loadings and default thresholds holds. In interpreting these tests note that a p-value of greater than 0.05 indicates that the null hypothesis cannot be rejected at the 95% confidence level. For both the S&P and the Moody's data, all three index models appear to be consistent with the available performance data. The quadratic and linear index models do not do significantly better jobs describing the observed data than the constant index models.

6 Conclusion

So far as we are aware, this paper offers the first comparative study of alternative methods of estimating default correlations from historical ratings performance data. The bias and precision of each estimator is assessed using Monte Carlo simulations of ratings performance datasets similar in size and grade construction to actual rating agency sources. Our findings suggest there would be significant benefit to shifting from the grade-by-grade method of moments estimators commonly used to a joint maximum likelihood methodology.

Our simulations of the method of moments (MM) estimator suggest adequate performance in estimating PDs, but reveal significant bias towards zero in estimates of factor loadings. The downward bias is most severe for higher quality grades. Our joint MLE methodology (MLE2) is similar to MM in that it allows factor loading to vary freely across grades, but shows minimal bias and reduced standard errors. MLE2 outperforms MM because it is able to utilize information on cross-grade default correlations.

If one is willing to impose the assumption of a common value for factor loadings across grades, then one can do better still. The gain in performance from the restricted estimator (MLE3) is not large for lower quality grades, but for higher quality grades the standard error on the factor loading can drop by roughly 40%. Of course, this restricted estimator would provide misleading results if factor loading did indeed vary across grades.

In applied work, an estimator that blends the characteristics of MLE2 and MLE3 can be used. This estimator allows for the possibility that higher quality obligors have systematically higher or lower factor loadings than lower quality obligors, while still capturing the benefits of imposing structure on the relationship between PDs and factor loadings. Instead of fixing a single common value for all factor loading as in MLE3, factor loadings are expressed as a simple parametric function of default probabilities. This approach permits greater flexibility in fitting data than MLE3, but affords greater efficiency than MLE2.

In addition to the aforementioned efficiency gains, MLE3 or a blended version of

MLE2 and MLE3 provides two practical advantages over the less restrictive estimators. First, by limiting the number of parameters that must be estimated, cross-bucket restrictions on factor loadings go a long way toward solving identification problem that arise when the number of obligors in a bucket is small or when defaults are infrequent. When very few defaults are observed in a bucket, estimating all the parameters of the more general default models becomes difficult or impossible. Such circumstances may arise, for example, when buckets consist of a large number of narrowly-defined rating grades.[5] Second, and perhaps more important, making factor loadings a (possibly constant) parametric function of default thresholds ensures that a bucket's factor loading can be calculated directly from its PD. This provides a natural means for assigning factor loadings to bank rating grades that straddle or fall between rating agency grades.

The application to Moody's and S&P ratings data presented in this paper demonstrates the feasibility of the maximum likelihood approach to estimating default correlations in a real world setting. Though our empirical results are not definitive, we find that relatively strong restrictions on the way factor loadings vary across rating grades are not inconsistent with observed data.

A broader conclusion to be drawn from this study concerns them importance of using information efficiently. For the foreseeable future, we will never have as much data on ratings performance as we would wish. Reasonable cross-grade parameter restrictions allow us to squeeze information out of our data as efficiently as passible. Research currently in progress seeks to expand the approach developed here to make use of information on ratings transition histories, rather than just default histories. This research should help to refine measurements of credit risk for higher quality rating grades, where defaults are rare but downgrades are common.

[5]See Appendix A for additional details on identification issues.

Grade	PD	Default Threshold	No. of Obligors
A	0.0015	-2.9677	400
B	0.0100	-2.3263	250
C	0.0500	-1.6449	100

Table 1: Characteristics of simulated rating grades

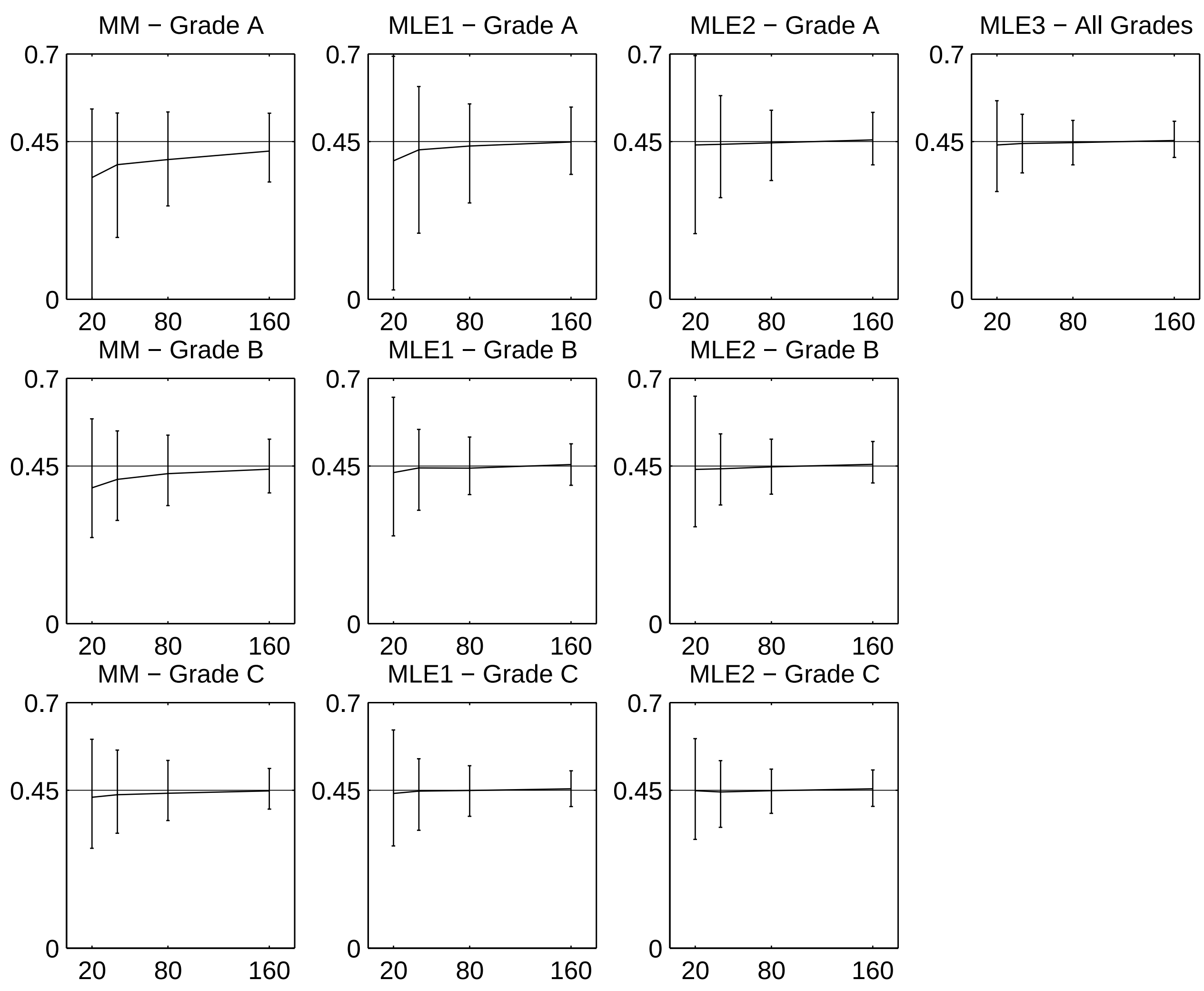

Figure 1: Median estimated factor loadings by sample size (error bars show 5th and 95th percentiles).

		MM			MLE1			MLE2			MLE3
T		A	B	C	A	B	C	A	B	C	All
	Mean	0.3251	0.3936	0.4303	0.3748	0.4272	0.4390	0.4356	0.4389	0.4427	0.4374
20	Std. Dev.	0.1435	0.0995	0.0917	0.1762	0.1053	0.0852	0.1319	0.0907	0.0773	0.0743
	RMSE	0.1901	0.1143	0.0937	0.1914	0.1077	0.0858	0.1325	0.0913	0.0775	0.0753
	Mean	0.3721	0.4166	0.4382	0.4151	0.4429	0.4462	0.4418	0.4426	0.4444	0.4454
40	Std. Dev.	0.1078	0.0792	0.0698	0.1269	0.0717	0.0609	0.0885	0.0610	0.0550	0.0529
	RMSE	0.1329	0.0859	0.0707	0.1315	0.0720	0.0610	0.0888	0.0614	0.0553	0.0530
	Mean	0.4003	0.4301	0.4439	0.4329	0.4468	0.4499	0.4455	0.4481	0.4497	0.4486
80	Std. Dev.	0.0779	0.0611	0.0498	0.0837	0.0512	0.0432	0.0619	0.0454	0.0401	0.0396
	RMSE	0.0923	0.0642	0.0501	0.0853	0.0513	0.0432	0.0620	0.0454	0.0400	0.0396
	Mean	0.4261	0.4427	0.4503	0.4527	0.4536	0.4548	0.4573	0.4563	0.4571	0.4548
160	Std. Dev.	0.0590	0.0468	0.0358	0.0595	0.0368	0.0320	0.0467	0.0369	0.0333	0.0311
	RMSE	0.0636	0.0474	0.0358	0.0595	0.0369	0.0324	0.0472	0.0374	0.0341	0.0315

Table 2: Distribution of estimated factor loadings by sample size for $w = 0.45$.

		MM			MLE1			MLE2			MLE3		
w		A	B	C	A	B	C	A	B	C	A	B	C
	Mean	-2.983	-2.331	-1.647	-2.982	-2.331	-1.647	-2.981	-2.331	-1.647	-2.982	-2.331	-1.647
0.15	Std. Dev.	0.096	0.063	0.060	0.096	0.063	0.060	0.097	0.063	0.060	0.096	0.063	0.060
	RMSE	0.097	0.063	0.060	0.097	0.063	0.060	0.097	0.064	0.060	0.097	0.063	0.060
	Mean	-2.992	-2.335	-1.651	-2.987	-2.334	-1.651	-2.985	-2.335	-1.653	-2.988	-2.334	-1.651
0.30	Std. Dev.	0.124	0.093	0.087	0.125	0.094	0.088	0.125	0.093	0.088	0.124	0.094	0.088
	RMSE	0.126	0.093	0.087	0.126	0.094	0.088	0.126	0.094	0.088	0.126	0.094	0.088
	Mean	-3.012	-2.349	-1.659	-3.008	-2.343	-1.657	-2.995	-2.345	-1.661	-3.005	-2.350	-1.664
0.45	Std. Dev.	0.175	0.138	0.124	0.173	0.137	0.123	0.163	0.127	0.117	0.165	0.132	0.120
	RMSE	0.180	0.140	0.124	0.177	0.138	0.124	0.165	0.129	0.118	0.170	0.134	0.122
	Mean	-3.042	-2.358	-1.653	-3.046	-2.360	-1.653	-3.009	-2.345	-1.652	-3.014	-2.360	-1.667
0.60	Std. Dev.	0.238	0.202	0.165	0.213	0.186	0.155	0.181	0.155	0.133	0.185	0.159	0.137
	RMSE	0.249	0.205	0.165	0.227	0.189	0.155	0.186	0.156	0.133	0.191	0.163	0.139
	True Value	-2.968	-2.326	-1.645	-2.968	-2.326	-1.645	-2.968	-2.326	-1.645	-2.968	-2.326	-1.645

Table 3(a): Distribution of estimated default thresholds by “true” factor loadings.

		MM			MLE1			MLE2			MLE3		
w		A	B	C	A	B	C	A	B	C	A	B	C
	Mean	0.149	1.000	5.008	0.149	1.000	5.008	0.150	0.999	5.008	0.149	1.000	5.008
0.15	Std. Dev.	0.045	0.166	0.616	0.045	0.167	0.616	0.045	0.167	0.618	0.045	0.167	0.620
	RMSE	0.045	0.166	0.616	0.045	0.166	0.615	0.045	0.167	0.618	0.045	0.167	0.619
	Mean	0.149	1.002	4.998	0.151	1.007	5.003	0.152	1.003	4.986	0.151	1.008	5.006
0.30	Std. Dev.	0.058	0.247	0.896	0.059	0.252	0.909	0.061	0.248	0.908	0.059	0.249	0.911
	RMSE	0.058	0.247	0.896	0.059	0.252	0.909	0.061	0.248	0.908	0.059	0.249	0.911
	Mean	0.150	0.999	4.985	0.152	1.014	5.006	0.155	1.000	4.953	0.151	0.990	4.923
0.45	Std. Dev.	0.095	0.379	1.286	0.089	0.381	1.264	0.085	0.338	1.198	0.083	0.351	1.222
	RMSE	0.095	0.379	1.286	0.089	0.381	1.263	0.085	0.338	1.198	0.083	0.351	1.224
	Mean	0.156	1.041	5.144	0.145	1.015	5.121	0.152	1.020	5.078	0.150	0.984	4.926
0.60	Std. Dev.	0.145	0.582	1.757	0.108	0.525	1.646	0.085	0.399	1.372	0.085	0.387	1.369
	RMSE	0.145	0.583	1.762	0.108	0.525	1.650	0.085	0.400	1.374	0.085	0.387	1.370
	True Value	0.150	1.000	5.000	0.150	1.000	5.000	0.150	1.000	5.000	0.150	1.000	5.000

Table 3(b): Distribution of estimated default probabilities by “true” factor loadings (in percentage points).

	MM			MLE1			MLE2			MLE3
	A	B	C	A	B	C	A	B	C	All
Mean	0.1161	0.1222	0.1321	0.1220	0.1180	0.1257	0.1643	0.1383	0.1412	0.1341
Std. Dev.	0.1146	0.0837	0.0761	0.1174	0.0813	0.0736	0.1032	0.0703	0.0638	0.0533
RMSE	0.1195	0.0882	0.0782	0.1206	0.0874	0.0775	0.1042	0.0712	0.0644	0.0556
Percentile										
2.5	0.0000	0.0000	0.0000	0.0013	0.0010	0.0019	0.0095	0.0099	0.0142	0.0111
5.0	0.0000	0.0000	0.0000	0.0021	0.0028	0.0042	0.0200	0.0194	0.0273	0.0318
50.0 (Med.)	0.0963	0.1338	0.1405	0.0955	0.1212	0.1314	0.1544	0.1379	0.1431	0.1387
95.0	0.3069	0.2495	0.2453	0.3390	0.2540	0.2430	0.3450	0.2570	0.2438	0.2173
97.5	0.3431	0.2736	0.2637	0.3814	0.2772	0.2622	0.3817	0.2793	0.2595	0.2305

Table 4(a): Distribution of estimated factor loadings for $w = 0.15$.

	MM			MLE1			MLE2			MLE3
	A	B	C	A	B	C	A	B	C	All
Mean	0.2172	0.2647	0.2800	0.2354	0.2723	0.2779	0.2898	0.2847	0.2850	0.2849
Std. Dev.	0.1306	0.0791	0.0752	0.1519	0.0863	0.0757	0.1173	0.0773	0.0707	0.0621
RMSE	0.1546	0.0866	0.0778	0.1650	0.0906	0.0788	0.1177	0.0788	0.0723	0.0639
Percentile										
2.5	0.0000	0.1034	0.1303	0.0040	0.0820	0.1179	0.0476	0.1307	0.1377	0.1605
5.0	0.0000	0.1344	0.1583	0.0081	0.1216	0.1498	0.0892	0.1529	0.1660	0.1793
50.0 (Med.)	0.2383	0.2667	0.2809	0.2393	0.2750	0.2813	0.2900	0.2841	0.2876	0.2871
95.0	0.4012	0.3934	0.3963	0.4784	0.4072	0.3968	0.4862	0.4115	0.3961	0.3858
97.5	0.4284	0.4168	0.4266	0.5429	0.4246	0.4199	0.5280	0.4333	0.4171	0.4018

Table 4(b): Distribution of estimated factor loadings for $w = 0.30$.

	MM			MLE1			MLE2			MLE3
	A	B	C	A	B	C	A	B	C	All
Mean	0.3158	0.3868	0.4150	0.3591	0.4209	0.4251	0.4289	0.4319	0.4278	0.4280
Std. Dev.	0.1457	0.0954	0.0919	0.1732	0.1022	0.0865	0.1255	0.0880	0.0796	0.0753
RMSE	0.1980	0.1144	0.0983	0.1955	0.1062	0.0900	0.1272	0.0898	0.0826	0.0784
Percentile										
2.5	0.0000	0.1998	0.2452	0.0119	0.2026	0.2553	0.1485	0.2475	0.2685	0.2813
5.0	0.0000	0.2333	0.2734	0.0238	0.2484	0.2838	0.2061	0.2816	0.2924	0.2986
50.0 (Med.)	0.3390	0.3849	0.4115	0.3849	0.4258	0.4277	0.4362	0.4354	0.4318	0.4309
95.0	0.5145	0.5520	0.5720	0.6184	0.5780	0.5598	0.6215	0.5677	0.5527	0.5495
97.5	0.5393	0.5906	0.6055	0.6493	0.6127	0.5788	0.6499	0.5968	0.5777	0.5702

Table 4(c): Distribution of estimated factor loadings for $w = 0.45$.

	MM			MLE1			MLE2			MLE3
	A	B	C	A	B	C	A	B	C	All
Mean	0.3856	0.4941	0.5566	0.4374	0.5517	0.5740	0.5384	0.5721	0.5767	0.5733
Std. Dev.	0.1686	0.1112	0.1038	0.2004	0.1095	0.0842	0.1193	0.0891	0.0749	0.0721
RMSE	0.2728	0.1535	0.1125	0.2580	0.1196	0.0881	0.1342	0.0933	0.0784	0.0768
Percentile										
2.5	0.0000	0.2947	0.3658	0.0107	0.3160	0.3930	0.2548	0.3755	0.4081	0.4004
5.0	0.0000	0.3219	0.3906	0.0248	0.3610	0.4267	0.3112	0.4122	0.4377	0.4368
50.0 (Med.)	0.4149	0.4896	0.5503	0.4769	0.5680	0.5747	0.5606	0.5843	0.5795	0.5822
95.0	0.6007	0.6852	0.7366	0.6803	0.7037	0.7020	0.6890	0.7008	0.6917	0.6742
97.5	0.6304	0.7267	0.7663	0.6970	0.7214	0.7294	0.7037	0.7244	0.7150	0.6895

Table 4(d): Distribution of estimated factor loadings for $w = 0.60$.

	Unrestricted Model	Quadratic Index	Linear Index	Constant Index
A	−3.3340	−3.3183	−3.3376	−3.3370
	(0.1562)	(0.1619)	(0.1577)	(0.1522)
BBB	−2.9179	−2.9033	−2.9193	−2.9196
	(0.1118)	(0.1192)	(0.1176)	(0.1119)
BB	−2.3309	−2.3296	−2.3400	−2.3394
	(0.0919)	(0.0842)	(0.0867)	(0.0820)
B	−1.6716	−1.6611	−1.6704	−1.6701
	(0.0599)	(0.0647)	(0.0685)	(0.0663)
CCC	−0.9361	−0.9252	−0.9337	−0.9351
	(0.0836)	(0.0818)	(0.0827)	(0.0828)

Table 5: Estimated default thresholds for S&P data.

	Unrestricted Model	Quadratic Index	Linear Index	Constant Index
A	−3.7697	−3.7717	−3.7654	−3.7731
	(0.2546)	(0.2558)	(0.2590)	(0.2547)
Baa	−3.0004	−3.0063	−3.0067	−3.0110
	(0.1026)	(0.1064)	(0.1127)	(0.1033)
Ba	−2.2418	−2.2518	−2.2534	−2.2488
	(0.0741)	(0.0676)	(0.0720)	(0.0676)
B	−1.5217	−1.5293	−1.5284	−1.5315
	(0.0550)	(0.0571)	(0.0595)	(0.0621)
Caa	−0.8561	−0.8743	−0.8690	−0.8618
	(0.0758)	(0.0734)	(0.0736)	(0.0793)

Table 6: Estimated default thresholds for Moody's data.

	Unrestricted Model	Quadratic Index	Linear Index	Constant Index
A	0.2754	0.2956	0.2568	0.2375
	(0.1748)	(0.1389)	(0.0849)	(0.0449)
BBB	0.2468	0.2729	0.2514	0.2375
	(0.1092)	(0.0884)	(0.0708)	(0.0449)
BB	0.2976	0.2500	0.2438	0.2375
	(0.0746)	(0.0546)	(0.0546)	(0.0449)
B	0.2223	0.2366	0.2349	0.2375
	(0.0478)	(0.0472)	(0.0466)	(0.0449)
CCC	0.2557	0.2395	0.2250	0.2375
	(0.0718)	(0.0690)	(0.0577)	(0.0449)

Table 5(a): Estimated factor loadings for S&P data.

	Unrestricted Model	Quadratic Index	Linear Index	Constant Index
A	0.2347	0.3236	0.3680	0.2807
	(0.2878)	(0.1936)	(0.0861)	(0.0417)
Baa	0.2895	0.3230	0.3359	0.2807
	(0.0994)	(0.0858)	(0.0661)	(0.0417)
Ba	0.3338	0.3063	0.3020	0.2807
	(0.0533)	(0.0455)	(0.0477)	(0.0417)
B	0.2592	0.2738	0.2676	0.2807
	(0.0400)	(0.0395)	(0.0396)	(0.0417)
Caa	0.2507	0.2287	0.2347	0.2807
	(0.0623)	(0.0588)	(0.0499)	(0.0417)

Table 5(b): Estimated factor loadings for Moody's data.

	Quadratic Index	Linear Index	Constant Index
Index Parameters			
Constant Term (β_0)	0.4488	0.3484	0.3815
	(0.3967)	(0.1581)	(0.0813)
Linear Term (β_1)	0.0867	−0.0241	
	(0.4278)	(0.0801)	
Quadratic Term (β_2)	0.0309		
	(0.1137)		
Specification Test vs. Unrestricted Model			
Likelihood-Ratio	1.4316	1.4462	1.9132
Restrictions	2	3	4
P-value	0.4888	0.6847	0.7512

Table 6(a): Index model parameter estimates for S&P data.

	Quadratic Index	Linear Index	Constant Index
Index Parameters			
Constant Term (β_0)	0.2261	0.3065	0.4718
	(0.3250)	(0.1418)	(0.0801)
Linear Term (β_1)	−0.1960	−0.0919	
	(0.3621)	(0.0799)	
Quadratic Term (β_2)	−0.0287		
	(0.0976)		
Specification Test vs. Unrestricted Model			
Likelihood-Ratio	1.7460	1.8172	3.1205
Restrictions	2	3	4
P-value	0.4177	0.6112	0.5379

Table 6(b): Index model parameter estimates for Moody's data.

Appendix

A Identification and Convergence Problems

In the main Monte Carlo study four sets of 1,500 synthetic datasets were constructed with w set to 0.15, 0.30, 0.45, and 0.60. For some of these datasets, one or more of the estimators described in Section 3 failed to generate a full set of model parameters. The table below shows the fraction of simulations for which one or more parameters could not be estimated.

w	MM	MLE1	MLE2	MLE3
0.15	0.005	0.000	0.000	0.000
0.30	0.005	0.003	0.003	0.000
0.45	0.007	0.038	0.043	0.007
0.60	0.061	0.281	0.311	0.121

For grades where the PD implied by γ_g is small, a simulated dataset may contain a very small number of defaults. This outcome is particularly likely when w is large. When fewer than two defaults are observed in a grade, the unrestricted model parameters are not identified. In essence, the data contain too little information to estimate separate default thresholds and factor loadings for each grade. Thus, MM and MLE1 cannot be used.

Even when model parameters are strictly identified by the data, the optimization algorithm used to obtain maximum likelihood estimators may fail to converge to a solution. Often such convergence problems arise when the matrix of second partial derivatives of the log-likelihood function (the Hessian matrix) is nearly singular. Work by Rothenberg (1971) shows that such singularity may result when model parameters are "nearly" unidentified. In general, highly correlated observations contain less information that is helpful in identifying model parameters than independent data. For this reason, it is perhaps not surprising that convergence problems are greater for higher values of w. Identification problems can be overcome by imposing parametric restrictions such as R3. This helps explain why MLE3 is more likely to converge to a solution than MLE1 or MLE2.

References

Brand, L. and Bahar, R.: 1999, Ratings performance 1999: Stability and transition, *Technical report*, Standard and Poor's.

Gordy, M. B.: 2000, A comparative anatomy of credit risk models, *Journal of Banking and Finance* **24**, 119–149.

Gordy, M. B.: 2001, A risk factor foundation for ratings-based bank capital rules. Working Paper.

Gupton, G., Finger, C. and Bhatia, M.: 1997, Creditmetrics – technical document, *Technical report*, The RiskMatrics Group.

Merton, R. C.: 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* **29**, 449–470.

Rothenberg, T. J.: 1971, Identification in parametric models, *Econometrica* **39**(3), 577–591.