Modeling stylized features in default rates

Emanuele Taufer

Abstract

We propose a stochastic model for the probability of default based on diffusions with given marginal distribution and autocorrelation function. The model tries to capture stylized features observed in historical default rates and is analytically tractable. Estimation procedures and expressions for analysis and prediction are provided.

AMS classifications: 62M10, 60G35.

Keywords: Default probability; diffusion process; beta distribution; Moody’s classes

1 Introduction

With the publication of the International Convergence of Capital Measurement and Capital Standards: a Revised Framework 2, the efforts of the Basel Committee on Banking Supervision (BCBS) to revise the standards governing the capital adequacy of internationally active banks achieved a critical milestone. Among other things, the "Basel II" framework, or Revised Framework, as the new standard is frequently called, is intended to promote a more forward-looking approach to capital supervision, one that encourages banks to identify the risks they may face, today and in the future, and to develop or improve their ability to manage those risks.

*Department of Computer and Management Sciences, University of Trento, Via Inama 5, 38100 Trento - Italy. emanuele.taufer@unitn.it, Ph. +39 0461 882368, Fax +39 0461 882124.
Practical applications of risk monitoring call for models which are intuitive and with light implementational burden. With this aim in focus, in this paper we will discuss a relatively simple model for default rates that seeks to capture most of the stylized features one observes in actual data while staying analytically tractable.

Recent contributions to this problem are due to Pederzoli and Torricelli (2005) and Marcucci e Quagliariello (2005) which discuss the effect of macroeconomic indicators over the defaults, Mira and Tenconi (2004) which use a logistic regression model, or Amerio et al. (2004) which provide an interesting approach based on Polya’s urn processes which, among other things, captures the marginal distribution of the defaults over a historical period. Another interesting contribution is that of Keenan et al. (1999) which provide a model based on the Poisson distribution with parameter depending upon macroeconomic indices. We refer the interested reader to the above mentioned papers and the references therein for further details on default rates modeling literature.

In our approach to the problem we do not start form a specification of economic fundamentals that may have effect on default risk. Instead, we specify directly a stochastic process for the default rate itself; the postulated process will encapsulate either the marginal and the correlation structure observed in empirical default rates. In our model, the stochastic process for the default rate exhibits mean reversion of the Ornstein-Uhlenbeck type and the instantaneous volatility depends on the level of the default rate and decreases as this approaches zero. As we will see, these empirical facts are well evident from historical data and support the implementation of a model which is able to obtain useful information, for prediction and analysis, from its past history. A further advantage will be given by the simple analytical forms of the conditional variances and expectations which can be exploited for estimation and prediction purposes.

The use of stochastic models with given autocorrelation and (non-Normal)
marginal structure is quite recent in the literature and find its earliest contributions in the papers of Barndorff-Nielsen (1998) and Barndorff-Nielsen and Shephard (2001), which exploit Ornstein-Uhlenbeck processes and self-decomposable distributions, Bibby et al. (2005), which exploit diffusion models. Both approaches try to produce flexible and analytically tractable models. In particular, the approach of Bibby et al. (2005) fits quite well in our context.

Our model is closely connected to the approaches and models of De Jong et al. (2001) which focus on exchange rates in a target zone and the celebrated one of Cox et al. (1985) for the term structure of interest rates.

In the next Section we will analyze past history of some empirical data on defaults and in Section 3 we will present a diffusion model with given marginal distribution. In Section 4 we will estimate parameters of our model for the data at hand and provide evidence on fit of the model.

2 Moody’s default rate statistics

Moody’s trailing 12-month default rates are widely monitored indicators of corporate credit quality and are a good source either for theoretical and empirical studies. For example, Amerio et al. (2005) have studied the historical distributions of one-year default rates for Ba-rated, B-rated and Caa-rated defaulters during the period 1970-1999; among other things they have found that the Beta distribution fits quite well all the classes considered; Keenan et al. (1999) have used either the entire Moody’s rated universe (all-corporate, AC) and a sub-grouping, i.e. the speculative-grade (SG) monthly data from 1970 to 1999 in order to provide a forecasting model.

In our study, in order to observe the phenomenon over a varied and long historical period, we are going to consider either AC and SG yearly data for the period
1920-2004. The data are taken from Moody’s Investor Service–Special Comment January 2005 and are freely available.

To begin with, we have a look at the linear plots of the two series in Figure 1. The paths of both series appear very similar, however, note that higher risks are, as expected, in the SG class. The series do not appear to be non-stationary, they however show periods of higher activity which show that volatility is not constant over the whole time horizon. The highest peak corresponds to the big crisis of the mid-thirties while, after 1970 there are more peaks.

**Figure 1.** Observed proportions of defaults for Moody’s All-Rated and Speculative–Grade classes. 1920-2004.

![Figure 1](image)

Let us now investigate the marginal and the dependence structure of these series. As far as the first of these two aspects is concerned, our yardstick is given by the Beta distribution with parameters α and β, i.e. a distribution having density

\[
f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1, \quad \alpha, \beta > 0,
\]

where \( B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta)^{-1} \) is the Beta function. Figure 2 depicts the PP plots of the Beta distribution for the two investigated series. As we see, for
this distribution, the fit is qualitatively good in both cases except for some years with no defaults which appear on the bottom of the graphs. Analogous results have been obtained by Amerio et al. (2004) for Ba-rated, B-rated and Caa-rated classes. The Beta distribution seems then to be a good candidate for modeling the marginal structure of defaults: appropriate choices of the parameters $\alpha$ and $\beta$ allows to obtain a large variety of shapes of the density; as a further support, we mention that this distribution is a very common choice in Bayesian statistics when modeling the law of a random probability over a given event.

**Figure 2.** Beta distribution P-P plots for All-Corporate and Speculative-Grade series.

![Beta distribution P-P plots](image)

By observing Figure 1 again, one can note that the actual range of default values is well below unity and hence ask whether the fit could be improved by choosing a distribution, such as the generalized Beta distribution, which can be made to assume values with positive probability in an interval $[0, b]$, with $b < 1$. Our computations show that there is no substantial difference between the two approaches if not for different values of estimated parameters. Moreover, recall that if $Y$ follows a generalized Beta distribution over $[0, b]$, then $X = Y/b$ follows a Beta over $[0,1]$, for further details about the generalized Beta and Beta distribution we
refer the reader to Johnson et al. (1994).

Next, we analyze the dependence structure of the defaults; we refer to Figure 3 where the autocorrelation and the partial autocorrelation functions are plotted for AC and SG.

**Figure 3.** Empirical ACF and Partial ACF for All-Corporate and Speculative-Grade with confidence limits

The autocorrelation function shows a constant positive decay for both series, indicating that an autoregressive process may be well suited for these data. The partial autocorrelation function shows a high spike in the first lag in both cases,
but for the AC series there is another significant negative spike at lag 3 which has however quite a low value. Both series show the same pattern indicating that information at time $t$, after conditioning at time $t - 1$, gains little from previous history of the process. Overall, the form of the autocorrelation and partial autocorrelation seems to indicate that a Markovian structure is appropriate for these series.

Notwithstanding the amount of data is not very large (85 observations for each series) their evidence is quite clear-cut, showing well definite structures and patterns that are quite remarkable for real data. Given these facts, it seems most appropriate to look for a model which tries to reproduce the empirical characteristics noted above. This will be the focus of the next section.

### 3 A diffusion model with given marginal distribution

In this section we will use a continuous stochastic process $\{X_t, t \geq 0\}$ as a generator of the flows of defaults over the period considered. The use of continuous models for discrete data, e.g. monthly or annual observations, is quite common given that discrete models may not be able to capture some features of the phenomenon at hand; indeed, there are authors that claim for its superiority over discrete models, some references and further discussion of direct interest here can be found in Lando and Skødeberg (2002) and Bladt and Sørensen (2005).

It is our aim to provide a stochastic model for $X_t$ which encapsulates the stylized features observed in the previous section. To do so we refer to Bibby et al. (2005) which have provided a general framework for construction of diffusion processes with given marginal distribution and autocorrelation function.

For our case, we will assume that the behavior of $X_t$ is governed by the following
stochastic differential equation:

\[ dX_t = -\lambda(X_t - \mu)dt + \sqrt{v(X_t)}dW_t, \quad t \geq 0, \] (3.1)

where \( \lambda > 0, \mu = E(X_t), v \) is a non-negative function and \( W_t \) is a standard Wiener process. If the distribution of \( X_0 \) is a Beta with parameters \( \alpha, \beta \) and

\[ v(x) = \frac{2\lambda}{\alpha + \beta} x(1-x) \] (3.2)

it follows at once from Theorem 2.3 in Bibby et al. (2005) that the diffusion process that solves (3.1) is Markovian and ergodic with invariant density (2.1) and autocorrelation function

\[ \text{Corr}(X_{t+h}, X_t) = e^{-\lambda h}. \] (3.3)

We also note that in such a case we have that

\[ E(X) = \mu = \frac{\alpha}{\alpha + \beta}, \quad V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \] (3.4)

Note that the proposed model includes all of the characteristics that have been pointed out by the empirical analysis over the AC and SG series: in particular a Beta marginal distribution and an exponentially decaying autocorrelation function. Moreover, the diffusion coefficient \( v(x) \) implies a non-constant volatility which decreases if the process \( X_t \) approaches its extremes. In actual cases the default rate \( X_t \) will stay low with a low value of the diffusion coefficient and the autoregressive part \( -\lambda(X_t - \mu) \) maintains the default rate close to its long run average. As \( X_t \) increases towards 0.5, the volatility will correspondingly increase mimicking what has been observed with the empirical analysis.

By using Îto formula we obtain the representation, for \( t, h \geq 0, \)

\[ X_{t+h} = e^{-\lambda h}X_t + \frac{\alpha}{\alpha + \beta} (1 - e^{-\lambda h}) + e^{-\lambda h} \int_t^{t+h} e^{\lambda(s-t)} \sqrt{v(X_s)}dW_s \] (3.5)
which can be used to obtain the conditional moments of the process; these may be helpful in forecasting.

As far as the conditional mean is concerned, we obtain immediately (see also Bibby et al. (2005)),
\[ E(X_{t+h}|X_t = x) = xe^{-\lambda h} + \frac{\alpha}{\alpha + \beta}(1 - e^{-\lambda h}). \]  
An expression for the conditional variance can be obtained with some more computations, note that, from (3.5) and (3.2) we can write
\[ Var(X_{t+h}|X_t = x) = e^{-2\lambda h} \frac{2\lambda}{\alpha + \beta} \int_{t}^{t+h} e^{2\lambda(s-t)} E[X_s (1 - X_s)|X_t = x] ds. \]  
For notational convenience set \( Var(X_{t+h}|X_t = x)e^{2\lambda h} = f(h) \), \( \alpha/(\alpha + \beta) = \mu \) and \( z = x - \mu \); then, deriving both terms of the above equation \( \text{wrt} \ h \), after a few computations, we obtain the following differential equation
\[ \frac{d}{dh}f(h) = - \frac{2\lambda}{\alpha + \beta} \left[ f(h) - \mu(1 - \mu)e^{2\lambda h} - z(1 - 2\mu)e^{\lambda h} + z^2 \right]. \]  
With the boundary condition \( f(0) = 0 \), this equation is solved by
\[ f(h) = (e^{-\frac{2\lambda}{\alpha + \beta}h} - 1)z^2 + (e^{\lambda h} - e^{-\frac{2\lambda}{\alpha + \beta}h}) \frac{2(\beta - \alpha)}{(\alpha + \beta)(\alpha + \beta + 2)} z^+ + (e^{2\lambda h} - e^{-\frac{2\lambda}{\alpha + \beta}h}) \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \]  
from which we can easily obtain back \( Var(X_{t+h}|X_t = x) \). Higher order conditional moments can be obtained in a similar way. As we see, the conditional variance depends on past values \( x \) and \( x^2 \); note also that the last coefficient on the right is the variance of the marginal Beta distribution.

\section{Fitting the model}

In order to fit the model to the data, we need to estimate the parameters \( \alpha, \beta \) and \( \lambda \). There is quite a bit of literature about estimation of discretely observed
continuous models, examples and theory can be found in Beskos et al. (2006), Larsen and Sørensen (2005), Bibby et al. (2004), Aït-Sahalia (2002), De Jong et al. (2001).

Here we partly follow the approach of Bibby et al. (2005) by splitting the problem in two parts: fitting the parameters of the marginal distribution on the one hand and the autoregression parameter on the other hand.

To estimate the marginal structure we need to estimate the parameters \( \alpha \) and \( \beta \) of the underlying Beta distribution. On the ground of quick applicability and simplicity, we will use a method of moments approach.

Let \( x_1, \ldots, x_T \) be the observed series of defaults and define the first two sample moments by \( \hat{\mu}_1 = T^{-1} \sum_{t=1}^{T} x_t \) and \( \hat{\mu}_2 = T^{-1} \sum_{t=1}^{T} x_t^2 \). Then the estimates of \( \alpha \) and \( \beta \) are quickly obtained as

\[
\hat{\alpha} = \frac{\hat{\mu}_1(\hat{\mu}_1 - \hat{\mu}_2)}{\hat{\mu}_2 - \hat{\mu}_1^2}, \quad \hat{\beta} = \frac{\hat{\alpha}(1 - \hat{\mu}_1)}{\hat{\mu}_1}.
\]

As far as the estimation of \( \lambda \) is concerned, we proceed by a least squares fitting of the empirical autocorrelation function, that is, we minimize, wrt \( \lambda \), the quadratic function

\[
f(\lambda) = \sum_{k=1}^{J} (\hat{\rho}_k - e^{-\lambda k})^2
\]

where \( \hat{\rho}_k \) denotes the lag \( k \) estimated autocorrelation. We tried different values of \( J \) obtaining estimates very close to each other. The values we report here are based on \( J = 14 \), a value which tries to fit the empirical autocorrelation function over a sufficiently large time span given the number of observations available.

The final estimates of the marginal and dependence parameters for the observed All-Corporate and Speculative-Grade series are provided in Table 1.

We mention that a simple fitting by \( J = 1 \), that is, just using the estimated autocorrelation at lag 1 obtains \( \lambda = 0.30 \) for AC and \( \lambda = 0.35 \) for SG.
Table 1. Estimates of $\alpha$, $\beta$ and $\lambda$ for All-Corporate and Speculative-Grade series.

<table>
<thead>
<tr>
<th></th>
<th>All- Rated</th>
<th>Speculative-Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.624</td>
<td>0.756</td>
</tr>
<tr>
<td>$\beta$</td>
<td>56.265</td>
<td>27.239</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.347</td>
<td>0.355</td>
</tr>
</tbody>
</table>

We use some graphical evidence to check how model (3.1) fits the data at hand. First of all, the PP plots of Figure 2 have been drawn by using the estimates of Table 1. As we have seen, the closeness of the historical distribution and the theoretical Beta is very good. These findings are sustained also by those of Amerio et al. (2004). As far as the correlation structure is concerned, Figure 4 reports the empirical autocorrelations and the theoretical one based on the estimated value of $\lambda$.

**Figure 4.** Empirical and fitted ACF for All-Corporate and Speculative grade.

As we see, in both cases, the closeness of the two curves is very good, especially as far as the decreasing speed of the first lags which may be quite important in analyzing mean reversion effects.

To validate further our findings we supplement the graphical model diagnostics by formal tests of hypothesis. We use a Generalized Likelihood Ratio (GLR) test in the form proposed by Fan and Zhang (2003) which has been found to be quite
powerful. Model (3.1) will be tested against a non parametric alternative of the form

\[ dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad (4.3) \]

where the functional form of drift and diffusion coefficients is left unspecified. The GLR is defined through the quantity

\[ l_T(h) = \frac{T - 1}{2} \log \frac{RSS_0}{RSS_1(h)}, \quad (4.4) \]

where \( h \) is a bandwidth, \( RSS_0 \) and \( RSS_1(h) \) are residuals sum of squares computed under the null hypothesis and the alternative hypothesis respectively. For more details about the testing procedure, as well as the construction of non parametric estimators for (4.3) we refer the interested reader to Fan and Zhang (2003).

As a starting point we test \( H_0 : dX_t = -\lambda(X_t - \mu)dt + \sigma(X_t)dW_t \), against model (4.3), i.e. linearity of the drift. In \( H_0 \) we use the estimates of \( \lambda, \alpha \) and \( \beta \) given in Table 1, while we use local linear regression (see, for example, Fan and Gijbels, 1996) to estimate \( \mu(X_t) \) and \( \sigma(X_t) \); in the computations of the non-parametric estimates we use the Epanechnikov kernel defined by \( K(u) = 0.75(1-u^2)I(|u| \leq 1) \), where \( I(\cdot) \) stands for the indicator function; as far as the choice of the bandwidth is concerned, we use \( h = s6T^{-2/9} \), where \( s \) is the standard deviation of \( x_1, \ldots, x_T \); this bandwidth is roughly viewed as "just right".

We now turn to numerical results. For the AC series we obtain \( l_T(h) = 16.49 \) and for the SG series the value is \( l_T(h) = 3.90 \). As far as critical points are concerned, we note that by heuristic arguments and Monte-Carlo simulations, Fan and Zhang (2003) argue that these do not depend sensitively on the true parameter values although they should depend on bandwidth and significance level. The values we have obtained for our test statistics are far beyond the critical values indicated in Table 3 of Fan and Zhang (2003) hence there is no evidence of departure from
the null hypothesis. As a further check, we set up a bootstrap procedure which provides, in all the above cases, bootstrap $p$-values which show again that there is no evidence against the null hypothesis.

Next, we test model (3.1) against model (4.3) obtaining the values $l_T(h) = 41.31$ for the AC series and $l_T(h) = 33.87$ for the SG series. Although values are larger than in the previous test, the values obtained are well beyond the critical values provided by Fan and Zhang (2003), again, bootstrap $p$-values do not show evidence against the null model.

5 Discussion

We have observed some of the empirical features of Moody’s default AC and SG series from 1920 to 2004. The series appears to be stationary, with non constant volatility and autoregressive features of the first order, moreover the historical distribution is fitted quite well by a Beta distribution.

Starting from these consideration we propose to model the behavior of the series by a diffusion process with given marginal distribution and autocorrelation function. The model is relatively simple and analytically tractable. Its intuitive interpretation may be appealing also to practitioners which may wish to analyze dynamics and structures of defaults based on freely available data. Also with this in mind we have fitted the model with simple estimation methods which can be implemented easily with standard computer packages. The main aim of the model is to help in understanding dynamics and relations of the phenomenon examined.

We recall also that the transition density is not known explicitly, we have provided expressions for conditional mean and variance which will help in evaluating the system.
References


Nr. 4  L. Erzegovesi, *Capire la volatilità con il modello binomiale*, Luglio 1999.


