SHANNON ENTROPY AS A MEASURE OF CERTAINTY IN A BAYESIAN CALIBRATION FRAMEWORK WITH BIVARIATE BETA PRIORS

L.J.S. Bodvin, A. Bekker and J.J.J. Roux
Department of Statistics, University of Pretoria, Pretoria, South Africa
E-mail: andriette.bekker@up.ac.za

Key words: Bayes estimation; bivariate beta distributions; calibration; credit ratings; multinomial distribution; probability of default; Shannon entropy.

Summary: The Bayesian estimator of the Shannon entropy is derived using Connor and Mosimann bivariate beta, bivariate beta type III and bivariate beta type V distribution distributions. Given the increased focus on the calculation of regulatory capital held by banks, it is important to have accurate probability of default estimates. Therefore in this paper the use of the Bayesian estimator of the Shannon entropy as a measure of certainty, when selecting the parameters of these various bivariate beta prior distributions in a Bayesian calibration framework, is illustrated using Moody’s corporate default rates.

1. Introduction

Having just survived what is arguably the worst financial crisis of our time, it is expected that the focus on regulatory capital held by financial institutions such as banks will increase significantly over the next few years. The probability of default is an important determinant of the amount of regulatory capital to be held, and the accurate calibration of this measure is vital. In this paper

AMS: 62F15, 62E
the use of the Shannon entropy when determining the parameters of a prior bivariate beta distribution as part of a Bayesian calibration methodology is illustrated. Various bivariate beta distributions will be considered as priors to the multinomial distribution associated with rating categories, and the appropriateness of these bivariate beta distributions will be tested on Moody’s default rate data. The exact expressions derived for the Bayesian estimation of Shannon entropy will be used to measure the certainty obtained when selecting the prior parameters.

This paper assumes a discrete random vector $X$ with multinomial distribution of dimension 3 and parameters $p_1, p_2, p_3$ and $n$, i.e., $X \sim \text{MultN}(p_1, p_2, p_3)$. In the case of large samples, accurate estimates of the parameter values $p_1, p_2$ and $p_3$ can be obtained easily, for example, using maximum likelihood estimation. For small samples the challenge increases considerably. One way to improve results is to incorporate prior information, which is done using a Bayesian approach.

The Bayesian estimator of the Shannon entropy is derived using different bivariate beta prior distributions for this multinomial model. Shannon entropy is defined as $H_3 = -\sum_{i=1}^{3} p_i \ln p_i$ by Shannon (1948), for discrete variables that can take one of three possible values, and will be used to measure the level of certainty associated with each of these bivariate beta distributions. Shannon entropy indicates the extent to which observations are concentrated around a single point, and thus measures the certainty or uncertainty present in the random variable. This measure is used in a variety of applications, amongst many others ecology (Pielou, 1967), cryptography (Simion, 2000 and Stephanides, 2005) and data mining (see Giudici, 2003).
In Section 2 the Bayesian estimator of the Shannon entropy will first be studied using the bivariate beta distribution as defined by Connor and Mosimann (1969), followed by the bivariate beta type III (Ehlers et al., 2009) and bivariate beta type V as prior distributions (Ehlers et al., 2010), the latter two both allow for positive correlation. In Section 3 the appropriateness of these bivariate beta distributions will be tested on Moody’s default rate data and the expressions derived for the Shannon entropy will be used to illustrate the effects of the different bivariate beta priors considered in this paper. Lastly, concluding remarks follow in Section 4. Appendix A contains some known results used in this paper.

2. Bayesian estimation of Shannon entropy

For the frequency data \((x_1, x_2, x_3)\) generated from the \(MultN(p_1, p_2, p_3)\) distribution the likelihood function is given by

\[
f(x|p_1, p_2) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n-x_1-x_2}
\]

where \(x_1 + x_2 + x_3 = n\), \(0 < p_i < 1\) for \(i = 1, 2\), and \(0 < p_1 + p_2 < 1\).

Assuming the squared error loss function, the Bayesian estimator of the Shannon entropy is given by the posterior mean. In this section the Bayesian estimator of the Shannon entropy is studied using the bivariate beta distribution (as defined by Connor and Mosimann), the bivariate beta type III distribution, and the bivariate beta type V distribution.

2.1 Connor and Mosimann bivariate beta prior

Consider as a prior the bivariate beta distribution of Connor and Mosimann (1969), denoted by \(BBeta^{CM}(\pi_1, \pi_2, \pi_3, d)\) and defined as:
\begin{align*}
f(p_1, p_2) &= \frac{\Gamma(\pi_1 + d)\Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_1^{\pi_1-1} p_2^{\pi_2-1} (1 - p_1 - p_2)^{\pi_3-1} \\
&\quad \times (1 - p_1)^{d - \pi_2 - \pi_3} \quad (2)
\end{align*}

where \(0 < p_i < 1\) for \(i = 1, 2\), \(0 < p_1 + p_2 < 1\), and \(\pi_1, \pi_2, \pi_3, d > 0\). For this distribution the correlation between \(P_1\) and \(P_2\) can only be negative, but if it is extended to more than two variables the generalised covariance structure allows for positive correlation, see Connor and Mosimann (1969). If \(d = \pi_2 + \pi_3\), (2) reduces to the well known bivariate beta type I distribution with density function given by

\begin{align*}
f(p_1, p_2) &= \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} p_1^{\pi_1-1} p_2^{\pi_2-1} (1 - p_1 - p_2)^{\pi_3-1} \quad (3)
\end{align*}

where \(0 < p_i < 1\) for \(i = 1, 2\), \(0 < p_1 + p_2 < 1\), and \(\pi_1, \pi_2, \pi_3 > 0\).

The prior distribution in (2) can be written as

\begin{align*}
f(p_1, p_2) &= \frac{\Gamma(\pi_1 + d)\Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} \sum_{r=0}^{\infty} \binom{d - \pi_2 - \pi_3}{r} \\
&\quad \times (-1)^r p_1^{\pi_1+r-1} p_2^{\pi_2+r-1} (1 - p_1 - p_2)^{\pi_3-1} \quad (4)
\end{align*}

by using the binomial expansion.

Using Bayes’ theorem,

\[ f(p_1, p_2|x) = \frac{f(x|p_1, p_2)f(p_1, p_2)}{\int \int f(x|p_1, p_2)f(p_1, p_2) dp_1 dp_2}, \]

it follows from (1) and (4) that
\[ f(x|p_1, p_2) f(p_1, p_2) = \frac{n!}{x_1! x_2! x_3!} \frac{\Gamma(\pi_1 + d) \Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1) \Gamma(\pi_2) \Gamma(\pi_3) \Gamma(d)} \times \sum_{r=0}^{\infty} \left( \frac{d - \pi_2 - \pi_3}{r} \right) \times (-1)^r p_1^{\pi_1 + x_1 + r - 1} p_2^{\pi_2 + x_2 - 1} (1 - p_1 - p_2)^{\pi_3 + x_3 - 1}, \]

and

\[ \int \int f(x|p_1, p_2) f(p_1, p_2) dp_1 dp_2 = \frac{n!}{x_1! x_2! x_3!} \frac{\Gamma(\pi_1 + d) \Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1) \Gamma(\pi_2) \Gamma(\pi_3) \Gamma(d)} \times (-1)^r B(\pi_1 + x_1 + r, \pi_2 + x_2, \pi_3 + x_3) \]

by using Definition 1 (Appendix A) and \( B(\alpha_1, \alpha_2, \alpha_3) = \frac{\prod_{i=1}^{\alpha_1} \Gamma(\alpha_i)}{\Gamma\left( \sum_{i=1}^{\alpha_1} \alpha_i \right)} \) denotes the beta function.

Therefore the posterior distribution is then given by

\[ f(p_1, p_2|x) = K \sum_{r=0}^{\infty} \left( \frac{d - \pi_2 - \pi_3}{r} \right) (-1)^r p_1^{\pi_1 + x_1 + r - 1} p_2^{\pi_2 + x_2 - 1} \]

where
\[ K = \frac{\Gamma(\pi_2 + x_2 + \pi_3 + x_3)\Gamma(\pi_1 + x_1 + x_2 + x_3 + d)}{\Gamma(\pi_1 + x_1)\Gamma(\pi_2 + x_2)\Gamma(\pi_3 + x_3)\Gamma(x_2 + x_3 + d)} \]  \quad (6)

\[ 0 < p_i < 1 \text{ for } i = 1, 2, 0 < p_1 + p_2 < 1, \text{ and } \pi_1, \pi_2, \pi_3, d > 0. \]

From (5), the Bayesian estimator of the Shannon entropy under squared error loss using the Connor and Mosimann bivariate beta distribution as a prior (denoted by \( H_{CM}^{3} \)) is derived as:

\[
H_{CM}^{3} = E_{f(p_1, p_2|x)}[H_{CM}^{3}]
\]
\[
= -K \int_0^1 \int_0^{1-p_2} p_i \ln p_i
\]
\[
\times \sum_{r=0}^{\infty} \left( \frac{d - \pi_2 - \pi_3}{r} \right) (-1)^r p_1^{\pi_1 + x_1 + r-1} p_2^{\pi_2 + x_2 - 1}
\]
\[
\times (1 - p_1 - p_2)^{\pi_3 + x_3 - 1} dp_1 dp_2
\]
\[
= -K \sum_{i=1}^{3} I_i,
\]

where \( K \) is defined by (6),

\[
I_i = \int_0^1 \int_0^{1-p_2} p_i \ln p_i
\]
\[
\times \sum_{r=0}^{\infty} \left( \frac{d - \pi_2 - \pi_3}{r} \right) (-1)^r p_1^{\pi_1 + x_1 + r-1} p_2^{\pi_2 + x_2 - 1}
\]
\[
\times (1 - p_1 - p_2)^{\pi_3 + x_3 - 1} dp_1 dp_2,
\]

for \( i = 1, 2 \), and

\[
I_3 = \int_0^1 \int_0^{1-p_2} (1 - p_1 - p_2) \ln(1 - p_1 - p_2)
\]
\[
\times \sum_{r=0}^{\infty} \left( \frac{d - \pi_2 - \pi_3}{r} \right) (-1)^r p_1^{\pi_1 + x_1 + r-1} p_2^{\pi_2 + x_2 - 1}
\]
\[
\times (1 - p_1 - p_2)^{\pi_3 + x_3 - 1} dp_1 dp_2.
\]
The expression $I_1$ can be written as

$$I_1 = \sum_{r=0}^{\infty} \left( \frac{d-\pi_2-\pi_3}{r} \right) (-1)^r \int_0^1 \int_0^{1-p_2} \frac{\partial}{\partial \pi_1} p_1^\pi_1 x_1 + r \cdot p_2^{\pi_2 + x_2 - 1} \cdot (1 - p_1 - p_2)^{\pi_3 + x_3 - 1} dp_1 dp_2$$

since $\frac{d}{dx} a^x = a^x \ln a$. Changing the order of integration and differentiation:

$$I_1 = \sum_{r=0}^{\infty} \left( \frac{d-\pi_2-\pi_3}{r} \right) (-1)^r \frac{\partial}{\partial \pi_1} B(\pi_1 + x_1 + r + 1, \pi_2 + x_2, \pi_3 + x_3)$$

$$= \sum_{r=0}^{\infty} \left( \frac{d-\pi_2-\pi_3}{r} \right) (-1)^r B(\pi_1 + x_1 + r + 1, \pi_2 + x_2, \pi_3 + x_3)$$

$$\times \left\{ \psi(\pi_1 + x_1 + r + 1) - \psi(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3 + 1) \right\}$$

where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ denotes the polygamma function. The simplification for $I_2$ and $I_3$ follows similarly.

Then

$$\hat{H}_3^{CM} = -K \sum_{r=0}^{\infty} \left( \frac{d-\pi_2-\pi_3}{r} \right) (-1)^r \frac{\Gamma(\delta_1)\Gamma(\delta_2)\Gamma(\delta_3)}{\Gamma(\sum_{j=1}^{\delta_3} \delta_j + 1)}$$

$$\times \sum_{j=1}^{\delta_3} \delta_j \left( \psi(\delta_1 + 1) - \psi(\sum_{j=1}^{\delta_3} \delta_j + 1) \right)$$

where $\delta_1 = \pi + x_1 + r, \delta_2 = \pi_2 + x_2, \delta_3 = \pi_3 + x_3$ and $K$ is defined in (6).

Figure 1 compares the Shannon entropy values obtained when using the Bayesian estimator derived in (7) for various combinations of $\pi_1, \pi_2, \pi_3$ and $d$. The multinomial frequencies were assumed to be $x_1 = 1, x_2 = 2$, and $x_3 = 10$. Decreasing $\pi_1$ or $\pi_2$ reduces $\hat{H}_3^{CM}$, indicating less uncertainty while increasing $\pi_1$ or $\pi_2$ increases $\hat{H}_3^{CM}$, indicating more uncertainty. Decreasing $\pi_3$ increases $\hat{H}_3^{CM}$ indicating more uncertainty, and increasing $\pi_3$ decreases $\hat{H}_3^{CM}$ indicating less uncertainty in the distribution. Larger values of $d$ are associated with lower Shannon entropy values, indicating less uncertainty. In
summary, as the concentration in the distribution remains closer to small values of $P_1$ and $P_2$, $\hat{H}_3^{CM}$ stays lower, but as soon as the concentration moves away from these small values to some point along the line $p_1 + p_2 = 1$ the uncertainty increases.

![Figure 1. Bayesian estimates of Shannon entropy: Connor and Mosimann bivariate beta prior](image)

**2.2 Bivariate beta type III prior**

Consider as a prior the bivariate beta distribution type III, denoted by $BBeta^{III}(\pi_1, \pi_2, \pi_3, c)$ (Ehlers et al., 2009 and Cardeño et al., 2005):

$$f(p_1, p_2) = \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \pi_1^{\pi_1 - 1} p_1^{\pi_1} \pi_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} \pi_3^{\pi_3} \times [1 - (1 - c)p_1 - (1 - c)p_2]^{-(\pi_1 + \pi_2 + \pi_3)}$$

(8)
where \(0 < p_i < 1\) for \(i = 1, 2\), \(0 < p_1 + p_2 < 1\), and \(\pi_1, \pi_2, \pi_3, c > 0\).

Ehlers et al. (2009) illustrated the effect of the parameter \(c\) in that it allows for positive correlation between \(P_1\) and \(P_2\), which is not the case with the Connor and Mosimann bivariate beta distribution.

From (1) and (8) it follows that

\[
\frac{f(x|p_1, p_2) f(p_1, p_2)}{n! \Gamma(\pi_1 + \pi_2 + \pi_3)^c \pi_1 \pi_2} \times \sum_{r=0}^{\infty} \sum_{s=0}^{r} \left( -\left( \frac{\pi_1 + \pi_2 + \pi_3}{r} \right) \right) \left( \frac{r}{s} \right) (c - 1)^r p_1^{\pi_1 + s + r - 1} p_2^{\pi_2 + s + r - 1} \times (1 - p_1 - p_2)^{\pi_3 + s - 1}
\]

(9)

by using the binomial expansion for the expression of the prior. Also,

\[
\int_0^1 \int_0^{1-p_2} f(x|p_1, p_2) f(p_1, p_2) dp_1 dp_2
\]

\[
= \frac{n! \Gamma(\pi_1 + \pi_2 + \pi_3)^c \pi_1 \pi_2}{x_1! x_2! x_3! \Gamma(\pi_1) \Gamma(\pi_2) \Gamma(\pi_3)} \times \sum_{r=0}^{\infty} \sum_{s=0}^{r} \left( -\left( \frac{\pi_1 + \pi_2 + \pi_3}{r} \right) \right) \left( \frac{r}{s} \right) (c - 1)^r B(\pi_1 + x_1 + r - s, \pi_2 + x_2 + s, \pi_3 + x_3)
\]

\[
= \frac{n! \Gamma(\pi_1 + \pi_2 + \pi_3)^c \pi_1 \pi_2}{x_1! x_2! x_3! \Gamma(\pi_1) \Gamma(\pi_2) \Gamma(\pi_3)} \times \sum_{r=0}^{\infty} \sum_{s=0}^{r} \left( -\left( \frac{\pi_1 + \pi_2 + \pi_3}{r} \right) \right) \left( \frac{r}{s} \right) (c - 1)^r \frac{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)}{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)} \times \frac{\Gamma(\pi_1 + x_1)}{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)} \times 2F_1(\pi_1 + \pi_2 + \pi_3, \pi_3 + x_3; \pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3; \frac{c - 1}{c}).
\]
Definition 1 (Appendix A), with $2F_1(\cdot)$ the Gauss hypergeometric function and the $B(\cdot)$ the beta function.

Thus the posterior distribution is given by

$$f(p_1, p_2 | x) = K \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \binom{-(\pi_1 + \pi_2 + \pi_3)}{r-s} \binom{r}{s} (c-1)^r p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1} (1 - p_1 - p_2)^{\pi_3+x_3-1}$$

where

$$K = \frac{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)}{\Gamma(\pi_1 + x_1) \Gamma(\pi_2 + x_2) \Gamma(\pi_3 + x_3)} c^{(\pi_1 + \pi_2 + \pi_3)}$$

$$\times \binom{2F_1(\pi_1 + \pi_2 + \pi_3, \pi_3 + x_3; \pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3; \frac{c-1}{c})}{\pi_1 + \pi_2 + \pi_3}$$

$$0 < p_i < 1 \text{ for } i = 1, 2, 0 < p_1 + p_2 < 1, \text{ and } \pi_1, \pi_2, \pi_3, c > 0.$$

Using the above, the Bayesian estimator of the Shannon entropy under squared error loss using the bivariate beta type III prior distribution (denoted by $\hat{H}_{III}$) is derived as:

$$\hat{H}_{III} = K \sum_{i=1}^{3} I_i$$

where $K$ is defined in (11),

$$I_i = \int_0^{1} \int_0^{1-p_2} p_i \ln p_i$$

$$\times \sum_{r=0}^{\infty} \sum_{s=0}^{r} \binom{-(\pi_1 + \pi_2 + \pi_3)}{r-s} \binom{r}{s} (c-1)^r p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1}$$

$$\times (1 - p_1 - p_2)^{\pi_3+x_3-1} dp_1 dp_2$$
for $i = 1, 2,$ and

$$I_3 = \int_0^1 \int_0^{1-p_2} (1 - p_1 - p_2) \ln(1 - p_1 - p_2)$$

$$\times \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \left(\frac{-\pi_1 + \pi_2 + \pi_3}{r}\right) \left(\frac{r}{s}\right) (c-1)^r p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1}$$

$$\times (1 - p_1 - p_2)^{\pi_3+x_3-1} dp_1 dp_2.$$ 

The expression $I_1$ can be simplified as follows

$$I = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \left(\frac{-\pi_1 + \pi_2 + \pi_3}{r}\right) \left(\frac{r}{s}\right) (c-1)^r$$

$$\times \int_0^1 \int_0^{1-p_2} \left[\frac{\partial^r}{\partial p_1^r} p_1^{\pi_1+x_1+r-s} \right] p_2^{\pi_2+x_2+s-1}$$

$$\times (1 - p_1 - p_2)^{\pi_3+x_3-1} dp_1 dp_2$$

$$= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \left(\frac{-\pi_1 + \pi_2 + \pi_3}{r}\right) \left(\frac{r}{s}\right)$$

$$\times (c-1)^r B(\pi_1 + x_1 + r - s + 1, \pi_2 + x_2 + s, \pi_3 + x_3)$$

$$\times [\psi(\pi_1 + x_1 + r - s + 1) - \psi(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3 + r + 1)]$$

with $\psi(\cdot)$ the polygamma function. Expressions for $I_2$ and $I_3$ follow similarly and the Bayesian estimator of the Shannon entropy for this case is:

$$\hat{H}_{3}^{III} = -K \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \left(\frac{-\pi_1 + \pi_2 + \pi_3}{r}\right) \left(\frac{r}{s}\right) (c-1)^r \frac{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)}{\Gamma(\gamma_1 + \gamma_2 + \gamma_3 + 1)}$$

$$\times \sum_{i=1}^{3} \gamma_i (\psi(\gamma_i + 1) - \psi(\sum_{j=1}^{3} \gamma_j + 1))$$

(12)
where $\gamma_1 = \pi + x_1 + r - s, \gamma_2 = \pi_2 + x_2 + s, \gamma_3 = \pi_3 + x_3$ and $K$ is defined in (11).

Figure 2 shows the Bayesian estimates of Shannon entropy values for various values of $c$, with $\pi_1 = \pi_2 = \pi_3 = 2$ and multinomial frequencies $x_1 = 1, x_2 = 2, \text{and } x_3 = 10$. Larger values of $c$ are associated with lower Shannon entropy values, indicating less uncertainty.

Figure 2. Bayesian estimates of Shannon entropy: bivariate beta type III prior
Remark

If $c = 1$, then (8) reduces to the bivariate beta type I distribution in (3) which is a conjugate prior for the multinomial distribution of dimension 3 defined in (1). Note that the correlation between $P_1$ and $P_2$ for the bivariate beta type I distribution can only be negative, see Balakrishnan and Lai (2009). The Bayesian estimator of the Shannon entropy using the bivariate beta type I prior under squared error loss is:

$$
\hat{H}_3 = -\sum_{i=1}^{3} \frac{\beta_i}{\sum_{j=1}^{3} \beta_j} (\psi(\beta_i + 1) - \psi(\sum_{j=1}^{3} \beta_j + 1))
$$

(13)

where $\beta_i = \pi_i + x_i$ for $i = 1, 2, 3$, with $\psi(x)$ the polygamma function. A generalization of this result can be found in Simion (1999).

2.3 Bivariate beta type V prior

Consider as a prior for the multinomial model in (1) the bivariate beta type V distribution, denoted as $BBeta^V(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c)$ and studied by Ehlers et al (2010) as:

$$
f(p_1, p_2) = \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \beta_1^{-\pi_1} \beta_2^{-\pi_2} c^{\pi_1 + \pi_2} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} \times (1 - p_1 - p_2)^{\pi_3 - 1} \times [1 - (1 - \frac{c}{\pi_1})p_1 - (1 - \frac{c}{\pi_2})p_2]^{-(\pi_1 + \pi_2 + \pi_3)}
$$

(14)

where $0 < p_i < 1$, for $i = 1, 2, 0 < p_1 + p_2 < 1$, and $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0$. If $\beta_1 = \beta_2 = 1$ then (14) reduces to the bivariate beta type III distribution in (8), and, if in addition $c = 1$, this reduces further to the bivariate beta type I distribution in (3). The inclusion of the additional parameters $\beta_1$ and $\beta_2$ also
add to the flexibility of the distribution, and increases the range over which positive correlation can be obtained.

In this case

\[
\int_0^1 \int_0^{1-p_2} f(p_1, p_2) f(x|p_1, p_2) dp_1 dp_2
\]

\[
= \frac{n!}{x_1! x_2! x_3!} \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1) \Gamma(\pi_2) \Gamma(\pi_3)} \beta_1^{-\pi_1} \beta_2^{-\pi_2} c^{\pi_1 + \pi_2} 
\]

\[
\times \frac{\Gamma(\pi_1 + x_1) \Gamma(\pi_2 + x_2) \Gamma(\pi_3 + x_3)}{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)}
\]

\[
\times F_1(\pi_1 + \pi_2 + \pi_3, \pi_1 + x_1, \pi_2 + x_2, \pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3; 1 - \frac{c}{\beta_1}, 1 - \frac{c}{\beta_2}).
\]

using (1), with \(F_1(\cdot)\) the hypergeometric function of two variables (see Definition 2, Appendix A). Then the posterior distribution, according to Bayes’ theorem, is given by

\[
f(p_1, p_2|x) = K p_1^{\pi_1 + x_1 - 1} p_2^{\pi_2 + x_2 - 1} (1 - p_1 - p_2)^{\pi_3 + x_3 - 1}
\]

\[
\times [1 - (1 - \frac{c}{\beta_1}) p_1 - (1 - \frac{c}{\beta_2}) p_2]^{-(\pi_1 + \pi_2 + \pi_3)}
\]

\[
= K \sum_0^\infty \sum_0^r \binom{r}{s} \frac{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)}{\Gamma(\pi_1 + x_1) \Gamma(\pi_2 + x_2) \Gamma(\pi_3 + x_3)}
\]

\[
\times p_1^{\pi_1 + x_1 + r - s - 1} p_2^{\pi_2 + x_2 + s - 1} (1 - p_1 - p_2)^{\pi_3 + x_3 - 1}
\]

(15)

where

\[
K = \frac{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)}{\Gamma(\pi_1 + x_1) \Gamma(\pi_2 + x_2) \Gamma(\pi_3 + x_3)}
\]

\[
\times \left[ F_1(\pi_1 + \pi_2 + \pi_3, \pi_1 + x_1, \pi_2 + x_2, \pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3; 1 - \frac{c}{\beta_1}, 1 - \frac{c}{\beta_2}) \right]^{-1},
\]

(16)

\(0 < p_i < 1\) for \(i = 1, 2\), \(0 < p_1 + p_2 < 1\), and \(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0\).
Following a similar approach as before, the Bayesian estimator of the Shannon entropy under squared error loss using the bivariate beta type V distribution as a prior (denoted by $H_3^V$) is derived as:

$$H_3^V = -K \sum_{r=0}^{\infty} \sum_{s=0}^{r} \binom{r}{s} \frac{(- (\pi_1 + \pi_2 + \pi_3)) (r)}{(r_1 - 1)^{r-s} (r_2 - 1)^s}$$

$$\times \frac{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)}{\Gamma(\gamma_1 + \gamma_2 + \gamma_3 + 1)} \sum_{i=1}^{3} \gamma_i (\psi(\gamma_i + 1) - \psi(\sum_{j=1}^{3} \gamma_j + 1))$$

where $\gamma_1 = \pi_1 + x_1 + r - s$, $\gamma_2 = \pi_2 + x_2 + s$, $\gamma_3 = \pi_3 + x_3$, with $K$ defined as in (16).

Figure 3 plots the Shannon entropy values for various values of $\beta_1$, $\beta_2$ and $c$, with $\pi_1 = \pi_2 = \pi_3 = 2$ and multinomial frequencies $x_1 = 1$, $x_2 = 2$, and $x_3 = 10$. Decreasing $\beta_1$ or $\beta_2$ respectively reduces $H_3^V$, indicating less uncertainty. Conversely, increasing $\beta_1$ or $\beta_2$ respectively increases $H_3^V$, indicating more uncertainty in the distribution. Decreasing $\beta_1$ and $\beta_2$ simultaneously increases $H_3^V$, indicating more uncertainty, whilst increasing $\beta_1$ and $\beta_2$ simultaneously decreases $H_3^V$, indicating less uncertainty. In general, larger values of $c$ are associated with lower Shannon entropy values, indicating less uncertainty. In summary, a larger concentration around small values of $P_1$ and $P_2$ is associated with lower $H_3^V$, and as the concentration moves towards the line $p_1 + p_2 = 1$ $H_3^V$ increases.
3. Shannon entropy in credit risk

3.1 Calibration: an overview

As a start, two key concepts are distinguished: default rate and probability of default (PD). The default rate corresponds to the actual number of customers who have defaulted out of a particular population of customers, whereas the probability of default is the likelihood of a particular customer defaulting. In the calibration of credit risk models, the default rate is used to determine the probability of default associated with a particular customer.

When data are readily available it is relatively easy to estimate the probability of default, this is typically done using logistic regression. In the environment of small low-default portfolios, it is almost impossible to construct meaningful logistic regression-type models to directly predict the default rate. However, Bayesian methods provide a systematic way to incorporate prior information and estimate the probability of default in a more informed manner. This is particularly relevant in the context of credit risk models, where a Bayesian approach can provide a more nuanced understanding of the risk profile of a portfolio.
likelihood of default, since adequate default information is generally not available. Instead, the probability of default of a customer is obtained indirectly, by assigning a credit rating (irrespective of whether they defaulted or not) based on some regression model. A probability of default is then assigned to that specific rating through a model calibration process. It is intuitive that the likelihood of default of a customer is influenced by, amongst others, the state of the economic cycle. A substantial amount of research in this field takes this into account.

The simplest calibration approach is to fit a PD curve to the credit ratings in the calibration sample such that the average of the calibration sample PD is equal to the long-run average of the portfolio, see Truck and Rachev (2005). This moment matching approach is sometimes combined with expert judgement, where PDs and PD bands are expertly assigned to the rating classes, see Pluto and Tasche (2005). Both these approaches have the risk of not being an accurate representation of the risk in the portfolio.

Rating transition matrices are used by Schuermann and Hanson (2004) and Truck and Rachev (2005), with a particular focus on the last column of the transition matrix (i.e. the default column). This enables the incorporation of economic conditions on the PD estimates.

The “most prudent estimation” methodology is contributed by Pluto and Tasche (2005), where they use upper confidence bounds to obtain PD estimates to any desired degree of conservatism, based on the assumption that PDs are monotonic between rating classes (which is generally true). Van der Burgth (2008) and Tasche (2010) estimated the PD curve based on the discriminatory power of the underlying rating model, measured by the receiver operating characteristic (ROC) and cumulative accuracy profiles (CAP).
The calibration methodologies discussed thus far are not explicit Bayesian calibration methodologies. However, the Bayesian estimation of credit risk, for both the underlying credit rating model and the model calibration, occurs more and more often.

As part of the credit rating model development, Loffler et al. (2005) propose a Bayesian methodology where they use as prior information the coefficients from credit rating models from other data sets. They find that “Bayesian estimators are significantly more accurate than the straight logit estimator”.

Gossl (2005) considers the development of a credit portfolio model using a Bayesian approach and proposes the use of the joint distribution of PDs and systemic correlation between the assets in a portfolio as opposed to the use of their point estimators.

Finally, the starting point of this analysis is obtained from Kiefer (2009). He considers the binomial distribution as an indication of the likelihood of a default or non-default event in a portfolio, and uses a univariate beta type I distribution as a prior for the binomial distribution. The parameters of his beta distribution were obtained by eliciting information from an expert, where the expert provided his/her opinion of the values to which the quantiles of the beta distribution should correspond to. Kiefer then used the method of moments to determine the beta distribution that satisfies the expert’s opinion. The univariate beta type I distribution was in turn used for calibration and to obtain confidence intervals for the PDs associated with the portfolio.

3.2 Setting

The aim of this analysis is to illustrate how the Shannon entropy can be used as a tool to select the prior distribution as part of a Bayesian credit risk model calibration approach. Following suit that probabilities of default are influenced
by the economic cycle, the differentiation between favourable and adverse economic conditions is considered.

Default rates differ between good and bad economic conditions, which is an aspect thoroughly investigated for macroeconomic stress testing purposes. The default rates in Figure 4 are taken from the Moody’s Annual Corporate Default Rate Study (2009), and represents the default rate out of their total rated (“All Rated”) population for each year. The US GDP values are the seasonally adjusted year-on-year GDP growth rates obtained from the US Bureau of Economic Analysis (BEA). The values are standardized in order to ease graphical interpretation.

**Figure 4**: Defaults rates over time
In favourable economic conditions, when a high GDP (dotted line) prevails, default rates (solid line) are low, and similarly, adverse economic conditions are associated with high default rates.

The default rates of the overall portfolio (“All Rated”) can be divided into two classes, namely investment grade and speculative grade. This model considers the following three events: (1) default occurring in investment grade, (2) default occurring in speculative grade, and (3) default does not occur.

If a default occurs, we are concerned with its rating quality, since a default in the investment grade class is more likely to have a larger impact on the bank’s book than a default in the speculative grade class. If a default does not occur, it is not really of interest since this is what the bank expects from the customer.

It is assumed that these three events follow the multinomial model in (1) and that the parameters of this model follow a bivariate beta distribution (discussed in Section 3.3). Given the practical importance of where defaults occur, it is clear that the use of the most appropriate bivariate beta distribution is very important.

3.3 Data

The default rate data used for this analysis is obtained from the Moody’s “Corporate Defaults and Recovery Rates, 1920-2008” study (2009) and spans from 1930 to 2008. The first 10 years from 1920 to 1929 are not used, but the rest of the data is used as-is, without making any assumptions regarding the quality of the data. Two sub-samples are selected to represent favourable and adverse economic conditions, i.e., a “Good” sample and a “Bad” sample. The “Good” sample consists of years where the GDP growth rate is larger
than the 60th percentile of the GDP distribution spanning the same period. Similarly, the “Bad” sample consists of years where the GDP growth rate is less than the 40th percentile of the GDP distribution. Observations with a GDP growth rate between the 40th and 60th percentiles are not used in this analysis in order to clearly illustrate the differences between favourable and adverse economic conditions. In practice, it is recommended to use all available data. Both datasets consist of 32 observations each.

For each sample, the investment and speculative grade default rates are used. The bivariate beta distributions investigated in this paper will be considered as priors to the joint distribution of the investment and speculative grade default events, as described above. In the data, there are quite a few years in which no defaults occurred. Theoretically this violates the assumption of the bivariate beta distributions that $p_i > 0$ for $i = 1, 2$. However, it is not believed that one should calibrate to a default rate of 0. This is an advantage of considering the bivariate beta distributions as priors, in that the distributions will be able to provide non-zero calibrated probabilities of default.

3.4 Analysis

Figure 5 compares the univariate distributions of the default rates for each of the categories between favourable and adverse economic conditions. Adverse economic conditions indicate less concentration in the low default rate region and the impact is particularly clear for the speculative grade default rates.
It is also expected that the likelihood of defaults occurring in either rating class increases in bad times and decreases in good times, indicating positive correlation between the two rating categories. The observed linear correlation is 0.643 in favourable economic conditions and 0.461 in adverse economic conditions. The positive correlation between investment grade and speculative grade default rates already indicates that the bivariate beta type III or bivariate beta type V distributions may be more appropriate as prior for this model.

Figure 6 compares the joint distributions and contour plots of default rates for the two samples.
3.5 Prior parameter selection

3.5.1 Traditional methods

Determining the parameters of the bivariate beta prior distributions with such little data proved to be quite challenging, as is generally the case with small samples. For the method of moments, credit experts can assign values to, say, the median, standard deviation, 5th percentile, 95th percentile, minimum, maximum, etc., see Kiefer (2009). This becomes quite difficult for more than one variable, in particular for credit experts who do not necessarily have a statistical background.

Figure 6. Joint distribution of investment and speculative grade default rates
Using maximum likelihood estimation (MLE), the likelihood function has to be optimised numerically since explicit expressions for each of the parameters cannot be obtained. An additional problem arising with MLE is that the parameter estimates are easily influenced by the observations.

Table 1 lists the parameter estimates obtained for the default rate distributions. Note that for the bivariate beta type V distribution the maximum likelihood estimates of the prior distribution parameters did not converge (indicated in the table as “DNC”). It is possible that this is due to the combination of the large number of parameters (6 parameters) to be estimated in conjunction with the small sample size (32 observations). The parameters that could be obtained with MLE could not capture the positive correlation observed.
Table 1. Parameter estimates: maximum likelihood estimation

<table>
<thead>
<tr>
<th>Economic Conditions</th>
<th>Parameter Estimates</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
<th>Type 6</th>
<th>Type 7</th>
<th>Type 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favourable</td>
<td>$\pi_1$</td>
<td>0.383</td>
<td>0.166</td>
<td>0.454</td>
<td>0.453</td>
<td>0.268</td>
<td>0.268</td>
<td>0.388</td>
<td>0.388</td>
</tr>
<tr>
<td></td>
<td>$\pi_2$</td>
<td>0.389</td>
<td>0.398</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td>$\pi_3$</td>
<td>0.399</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td>$\pi_4$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\pi_5$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\pi_6$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\pi_7$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\pi_8$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>0.359</td>
<td>0.457</td>
<td>0.379</td>
<td>0.246</td>
<td>0.270</td>
<td>0.270</td>
<td>0.364</td>
<td>0.364</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\beta_4$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\beta_5$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\beta_6$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\beta_7$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$\beta_8$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>$c_1$</td>
<td>76.097</td>
<td>59.421</td>
<td>76.155</td>
<td>76.155</td>
<td>76.155</td>
<td>76.155</td>
<td>76.155</td>
<td>76.155</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
<td>53.097</td>
<td>40.026</td>
<td>53.097</td>
<td>53.097</td>
<td>53.097</td>
<td>53.097</td>
<td>53.097</td>
<td>53.097</td>
</tr>
<tr>
<td></td>
<td>$c_3$</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
</tr>
<tr>
<td></td>
<td>$c_4$</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
</tr>
<tr>
<td></td>
<td>$c_5$</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
</tr>
<tr>
<td></td>
<td>$c_6$</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
</tr>
<tr>
<td></td>
<td>$c_7$</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
</tr>
<tr>
<td></td>
<td>$c_8$</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
<td>34.181</td>
</tr>
<tr>
<td></td>
<td>$\delta_1$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_2$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_3$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_4$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_5$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_6$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_7$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\delta_8$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_1$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_3$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_4$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_5$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_6$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_7$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>$\phi_8$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>Loglikelihood</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
3.5.2 Bayesian estimation of Shannon entropy

The proposal is to use the Shannon entropy to determine the optimal values of the parameters for the various bivariate beta priors considered for the multinomial model of dimension 3 in conjunction with the data available and expert judgement. In the financial sector Bayesian priors are mostly constructed using expert opinion. However, for illustrative purposes in this paper, the prior will be constructed using default data. In this application, only the bivariate beta type III and bivariate beta type V distributions are considered as possible priors due to their ability to account for positive correlation.

The following steps are used to determine the parameters:

1. Determine the order of magnitude of the parameters using the conclusions from the shape analyses conducted for the various bivariate beta distributions, see Bodvin (2010).
   
   (a) Favourable economic conditions: From the joint distributions in Figure 6 it is noted that for favourable economic conditions, the concentration of the distribution is towards small values of investment grade default rates ($P_1$) and small values of speculative grade default rates ($P_2$). For the bivariate beta type III distribution, this suggests a choice of parameters where $\pi_1, \pi_2$ and $\pi_3$ are less than $c$. For the bivariate beta type V distribution, this suggests a possible choice of parameters where $\pi_1, \pi_2, \pi_3, \beta_1$ and $\beta_2$ are less than $c$.

   (b) Adverse economic conditions: From the joint distributions in Figure 6 it is noted that for adverse economic conditions, the concentration of the distribution is towards larger values of the speculative grade default rates ($P_2$). For the bivariate beta type III distribution this can be obtained by choosing $\pi_1$ to be less than $\pi_2, \pi_3$ and $c$. For the bivariate beta type V distribution this can be obtained by choosing $\beta_1$ to be less than $\pi_1, \pi_2, \pi_3, \beta_2$ and $c$.

2. Determine bands for the parameters using quantitative analyst expert judgement (and trial and error). For example, for the bivariate beta type III distribution, $c$ has to be much larger than $\pi_1, \pi_2$ and $\pi_3$ to obtain the desired concentration.

3. Using a grid search approach and an arbitrary step size, calculate
the Shannon entropy using the Bayesian estimates in (12) and (17) respectively, with inputs:

(a) Bivariate beta distribution parameters: The combination of parameters in the grid.

(b) Multinomial distribution parameters: Since the focus of this analysis is on the selection of the prior distribution, \( x_1 = 1, x_2 = 2 \) and \( x_3 = 10 \) were used as the multinomial distribution observations. These can of course be changed as well.

4. Calculate the correlation for each combination of the parameters in the grid.

5. When selecting the parameters of the prior distributions, choose them such that:

(a) The parameters provide a Shannon entropy in a pre-specified range, keeping in mind that lower Shannon entropy values are associated with less uncertainty and therefore higher concentration in the distribution towards smaller values of \( P_1 \) and \( P_2 \). In this analysis, the bands are different for the favourable and adverse economic conditions, since the concentration in the observed distributions are different. Selecting the range of Shannon entropy can be done by trial and error. For this analysis, Shannon entropy was chosen to be between 0.35 and 0.45 for favourable economic conditions and between 0.45 and 0.55 for adverse economic conditions.

(b) The parameters provide a correlation similar to the observed correlation (0.64 for favourable economic conditions and 0.46 for adverse economic conditions). For this analysis, the correlation was chosen to be between 0.6 and 0.7 for favourable economic conditions, and between 0.4 and 0.5 for adverse economic conditions.

Table 2 and Figure 7 summarise the grid search results for the bivariate beta type III distribution. The first three columns provide information regarding the bounds used. The last two columns provide the parameters chosen for the two bivariate beta type III distributions. The parameters seem to reflect the characteristics of the observed distributions.
Table 2. Parameter selection: bivariate beta type III distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Step Size</th>
<th>Maximum</th>
<th>Favourable economic conditions</th>
<th>Adverse economic conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>$\pi_3$</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$c$</td>
<td>20</td>
<td>20</td>
<td>100</td>
<td>100</td>
<td>40</td>
</tr>
</tbody>
</table>

Shannon entropy 0.446 0.502
Correlation 0.698 0.491

Figure 7. Bivariate beta type III fitted distributions

Similarly, Table 3 and Figure 8 summarise the grid search results for the bivariate beta type V distribution. The first three columns provide information regarding the bounds used. The last two columns provide the parameters chosen for the two bivariate beta type V distributions. Again, the parameters reflect the characteristics of the observed distributions.
Table 3. Parameter selection: bivariate beta type V distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Step Size</th>
<th>Maximum</th>
<th>Favourable economic conditions</th>
<th>Adverse economic conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_1 )</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>( \pi_2 )</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>( \pi_3 )</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>( c )</td>
<td>20</td>
<td>20</td>
<td>100</td>
<td>80</td>
<td>40</td>
</tr>
</tbody>
</table>

| Shannon entropy | 0.395 | 0.488 |
| Correlation     | 0.624 | 0.423 |

Figure 8. Bivariate beta type V fitted distributions

Note that, in order to illustrate the results clearly, a very coarse grid has been used. In practice, it is advised to use a finer grid as this may significantly improve the accuracy of the parameter estimates. These results indicate that the bivariate beta type III and bivariate beta type V distributions are very flexible since they have the ability to deal with positive correlation in the underlying data. Note that for this example the general level of the Shannon entropy for the bivariate beta type V distribution is lower than for the bivariate beta type III distribution. This could possibly be as a result of the additional parameters.
Having more parameters implies that the distribution can be defined better, and therefore there is less uncertainty. For this example, the bivariate beta type III distribution appears to be the best candidate, since it only requires one additional parameter to take into account the positive correlation between the investment grade default rates and the speculative grade default rates. Using the Bayesian estimates of the Shannon entropy proved to be a useful aid in selecting the prior distribution when the sample size is small.

4. Conclusion

The flexibility resulting from using different bivariate beta distributions gives one the opportunity to include expert opinion and prior information to obtain more realistic results for a specific situation. Exact expressions for the Bayesian estimator of the Shannon entropy have been successfully derived for combinations of the multinomial distribution with various bivariate beta distributions as priors. The use of this estimator was illustrated by considering a Bayesian approach for the calibration of investment and speculative grade default rates, and proved to be a useful tool when selecting the appropriate parameters for the bivariate beta prior distributions.

Acknowledgement

The authors would like to thank the anonymous referee for his/her constructive comments and suggestions. This work is based upon research supported by the National Research Foundation, South Africa (GRANT: FA2007043000003).
References


Appendix A

Definitions and some lemmas which are used in this paper are presented in this appendix.

Definition 1. (Gradshteyn and Ryzhik, 2007, p. 1005)

The Gauss hypergeometric function is defined as

\[ 2F_1(\alpha, \beta; \gamma; x) = \sum_{k=0}^{\infty} \frac{(\alpha)_k (\beta)_k x^k}{(\gamma)_k k!} \]

where \((\alpha)_k = (\alpha+1)...(\alpha+k-1)\) and \((\alpha)_0 = 1\). The integral representation is

\[ 2F_1(\alpha, \beta; \gamma; x) = \frac{1}{B(\beta, \gamma - \beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tx)dt \]

for \(\text{Re}\, \gamma > 0\) and \(\text{Re}\, \beta > 0\).

Definition 2. (Gradshteyn and Ryzhik, 2007, p. 1018, 1021)

The hypergeometric function of two variables is defined as

\[ F_1(\alpha, \beta'; \gamma; x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(\alpha)_{m+n}(\beta')_m(\beta')_n}{(\gamma)_{m+n} m! n!} x^m y^n \]

for \(|x| < 1\) and \(|y| < 1\), and the integral representation is

\[ F_1(\alpha, \beta', \gamma; x, y) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\beta')\Gamma(\gamma - \beta - \beta')} \times \int \int \frac{u^{\beta-1}v^{\beta'-1}}{u + v \leq 1} (1-u-v)^{\gamma-\beta-\beta'-1}(1-ux-vy)^{-\alpha} du dv \]

for \(\text{Re}\, \beta > 0\), \(\text{Re}\, \beta' > 0\), and \(\text{Re}\, (\gamma - \beta - \beta') > 0\).
Relation 1. (Prudnikov et al., 1986, p. 566)

\[
\int \int_{u \geq 0, v \geq 0, u + v \leq 1} x^{\beta-1} y^{\beta'-1} (1 - x - y)^{\gamma - \beta - \beta' - 1} (1 - ux - vy)^{-a} \, dx \, dy \\
= \Gamma \left[ \begin{array}{c} \beta, \beta', \gamma - \beta - \beta' \\ \Gamma \end{array} \right] F_1(\alpha, \beta, \beta', \gamma; u, v)
\]

where \( \Gamma \left[ \begin{array}{c} a_1, a_2, \ldots, a_m \\ b_1, b_2, \ldots, b_n \end{array} \right] = \frac{\prod_{i=1}^{m} \Gamma(a_i)}{\prod_{j=1}^{n} \Gamma(b_j)} \), and \( F_1(\alpha, \beta, \beta', \gamma; u, v) \) is the hypergeometric function of two variables, see Definition 2.

Relation 2. (Gradshteyn and Ryzhik, 2007, p. 1008)

\[
\sum_{n=0}^{\infty} \frac{\Gamma(n+1)\Gamma(1+n\alpha)}{(n!)^2 \Gamma(n+1)} = \frac{2F_1(\alpha, \beta; \gamma; z)}{(1 - z)^{-\beta} 2F_1(\gamma - \alpha, \beta; \gamma; z = \frac{z}{z - 1})}.
\]

Relation 3. (Gradshteyn and Ryzhik, 2007, p. 1008)

\[
\sum_{n=1}^{\infty} \frac{\Gamma(n+1)\Gamma(n+n\alpha)}{(n!)^2 \Gamma(n+1)} = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}.
\]

Manuscript received, 2010.10.26, revised, 2011.04.25, accepted, 2011.05.21.