Nowadays, there is no doubt whatsoever about the decisive role of the business cycle and macroeconomic factors in measuring loss given default (LGD). By assuming two regimes characterizing the business cycle, expansion and recession, the LGD can be threat as a stochastic variable modelled through a mixture of an expansion and a recession distribution.

In this paper we are particularly interested in analyzing the accuracy of prediction intervals based on the widespread assumption of a single beta distribution when modelling recovery rates over the business cycle in the MKMV framework. Alternative prediction intervals are provided assuming mixtures of two beta distributions, distinguishing between secured and unsecured securities which are strongly or weakly sensitive to the business cycle. Additionally, we propose to construct bootstrap prediction intervals and we compare their performance in order to improve the coverage accuracy of prediction intervals.

**Keywords:** Business cycle, procyclicality, loss given default, Moody’s KMV model, finite mixtures of beta distributions, E-M algorithm, bootstrap prediction intervals.

Addresses for correspondence: Dr. Susana Alvarez, Department of Quantitative Methods for the Economy, University of Murcia (Spain), 30100 Campus Espinardo. E-mail: salvarez@um.es, Phone: +34968367902. Dr. J. Samuel Baixauli, Department of Management and Finance, University of Murcia (Spain), 30100 Campus Espinardo. E-mail: sbaixaul@um.es, Phone: +34968367819.
I. Introduction

The existence of the relationship between economic activity and credit risk measurement is not controversial (Casarin and Trecroci, 2006, and references in there). Borrowers’ risk varies with the state of the economy and hence risk-sensitive capital requirements are procyclical. Procyclicality is meant as the amplification of the business cycle provoked by the reduction in credit risk availability in recession periods (and vice versa in expansion periods). Banking is a procyclical business. Banks tend to contract their lending activity when business turn down and the reduced lending amplify the recession. In contrast, banks expand their lending activity during expansion periods, contributing to a possible overheating of the economy.

The business cycle and the main components of credit risk are linked (Koopman and Lucas, 2005). Its main components are the probability of default (PD), the loss given default (LGD) and the exposure at default (EAD) (Allen and Saunders, 2003; Lowe, 2002). The PD is the estimate of the likelihood of the borrower defaulting on its obligations within a given horizon; the LGD is defined as the loss incurred in the event of default and it is equal to one minus the recovery rate at default; and, the EAD is the nominal value of the borrower’s debt. The potential credit loss at a given horizon is computed as the product of the PD, the LGD and the EAD. Many different modelling techniques have been applied to determine the PD while much less attention has been paid to how estimate LGD (Shumway, 2001; Chava and Jarrow, 2004; Duffie et al., 2004). In the last years, regulatory bodies have become to focus more closely on LGD analysis since errors in estimating LGD are as important as errors in estimating the PD in determining potential credit losses. For instance, the advanced internal rating based (IRB) approach of Basel II leaves a bank quite a high amount of flexibility to determine the recovery rates for a loan, and this could be considered as a motivation for a bank use the more advanced IRB approach and provide an own sophisticated model for LGD.

The decisive role of the business cycle and macroeconomic factors in measuring LGD is absolutely accepted in the related literature (Altman et al., 2005; Carey, 1998; Pederzoli and Torricelli, 2005; Schürmann, 2004). On consequence, credit risk measurement models, as the Moody’s KMV model (MKMV), have begun to address the cyclicality in LGD.
In this paper we are particularly interested in analyzing the accuracy of prediction intervals based on the widespread assumption of a single beta distribution when modelling recovery rates over the business cycle in the MKMV framework. In so doing, we consider a dynamic behaviour of LGD over the business cycle. By assuming two regimes characterizing the business cycle, expansion and recession, we threat the LGD as a stochastic variable modelled through a mixture of an expansion and a recession distributions. The finite mixtures of beta distributions provide an extremely flexible method of modelling unknown distributional shapes over the business cycle, which apparently cannot be modelled by a single beta distribution function. The called EM (Expectation-Maximization) algorithm is used to fit the beta mixture models. We provide alternative prediction intervals assuming that the underlying distribution is a mixture of two beta distributions, distinguishing between secured and unsecured securities which are strongly or weakly sensitive to the business cycle. Additionally, we propose to construct bootstrap prediction intervals and we compare their performance in order to improve coverage accuracy of beta prediction intervals.

The remainder of the paper is organized as follows: Section II describes briefly how to obtain estimated prediction intervals in the MKMV framework. The finite mixtures of g component distributions and the EM algorithm are reviewed in Section III. Section IV presents a simulation study to analyze the performance of prediction intervals based on the assumption of a single beta distribution over the business cycle. Alternative prediction intervals are provided under the assumption of mixtures of two beta distributions as well as the percentile bootstrap and percentile-t bootstrap prediction intervals. In Section V we report the results of the simulation experiments to compare the six prediction intervals which were constructed in Section IV. A performance comparison between them is provided. Conclusions are drawn in Section VI.

II. LGD prediction intervals in the MKMV framework

Loss given default (LGD) is an essential factor to investors and lenders to estimate potential credit losses. Estimates of LGD have usually been done by traditional methodologies of historical averages segmented by debt type and seniority. In contrast
to traditional methodologies, some credit risk models, such as the Moody’s KMV model, pay attention to the prediction horizon, incorporating cyclic and firm specific effects. Moody’s researchers have found that the MKMV model is a much better predictor of LGD than the traditional methodology, helping institutions to better price and manage credit risk (Gupton and Stein, 2002, 2005). MKMV model gives estimates of LGD for defaults occurring immediately and at one year from the time of evaluation.

MKMV model is a statistical model that uses linear regression-based techniques to construct prediction intervals. It includes a number of explanatory variables, grouped into five categories: collateral and backing, debt type and seniority class, firm-level information, industry specific characteristics and, macroeconomic and geographic conditions. The dependent variable in MKMV model is the recovery rate in the event of default, defined as its market value approximately one-month after default. The resulting linear regression equation is the following:

\[
R_t = X_t' \beta + \epsilon_t, \ t = 1, \ldots, n
\]  

where, \( R_t = (1 - LGD) \) represents the recovery rate of a default instrument, \( X_t \) is the vector of explanatory variables and \( \epsilon_t \) denotes the error term of the model. MKMV model assumes that recovery rates are stochastic variables independent from default probabilities and drawn from a beta distribution, even though the business cycle might affect the shape of the underlying distribution.

If a single beta distribution is assumed to represent the behavior of recovery rates over the business cycle, recovery rates observed over different periods are interpreted as random variables drawn from a stable distribution. Alternatively, if two different distributions are considered to hold over expansion and recession periods, recovery rates can be seen as realizations of these two distributions. By positing two regimes, expansion and recession, recovery rates are drawn from a mixture of an expansion and a recession distribution. If the business cycle \( S \) states over the period \([t, t+k]\), recovery rates are random variables with the following state-dependent distribution:
\[
f(R_{t+k} | S) = \begin{cases} 
  f_E(R_{t+k}) & \text{if } S = E \quad \text{(Expansion)} \\
  f_R(R_{t+k}) & \text{if } S = R \quad \text{(Recession)}
\end{cases}
\] (2)

where, \( R_{t+k} \) = recovery rates over \([t, t+k]\), \( f_E(R_{t+k}) \) = probability density function of the recovery rate over \([t, t+k]\) conditional on the state of expansion, \( f_R(R_{t+k}) \) = probability density function of the recovery rate over \([t, t+k]\) conditional on the state of recession.

Assuming the regimes are unobserved, the probabilities of each state are \( P(E) = \pi \) and \( P(R) = (1 - \pi) \) and need to be estimated (Hamilton, 1989, Filardo, 1994). Hence, the prior density function of the recovery rate over the period \([t, t+k]\), \( f_i(R_{t+k}) \) is given by the following density function of a mixture of the two conditional beta distributions:

\[
f_i(R_{t+k}) = \pi f_E(R_{t+k}) + (1 - \pi) f_R(R_{t+k})
\] (3)

As MKMV assumes that a single beta distribution captures the observed behavior of recovery rates over the business cycle, in order to regress the recovery rates on the explanatory variables, this model converts the beta distributed recovery values to a more normally distributed dependent variable using the normal quantile transformation, \( \tilde{R}_i = N^{-1}(Beta(R_i, p, q)) \), where, \( N^{-1} \) is the inverse of the normal distribution function. As consequence, the linear regression equation is rewritten as follows:

\[
\tilde{R}_i = \beta_0 + \beta_1 X_i + v_i, \quad t = 1, \ldots, n
\] (4)

where, \( \tilde{R}_i \) is the transformed dependent variable and \( v_i \) is the error term.

However, the usual MKMV assumption of beta distribution over the business cycle could imply the fact of the recovery rates contained measurement error, defined as the difference between the observed value and the true value. That is, \( w_i = \tilde{R}_i - \tilde{R}_i^T \), where \( \tilde{R}_i^T = N^{-1}(F(R_i, \pi, p_1, q_1, p_2, q_2)) \), \( F(R_i, \pi, p_1, q_1, p_2, q_2) \) is the distribution function of a mixture of two beta distributions; \( p_1, q_1 \) are the parameters of the first beta distribution; \( p_2, q_2 \) are the parameters of the second beta distribution; and \( \pi \) is the
probability of each state. Using $\tilde{R}_i^T$, the linear regression equation is

$$\tilde{R}_i = \tilde{R}_i^T + w_i = (\beta_0 + \beta_1 X_i + u_i) + w_i = \beta_0 + \beta_1 X_i + (u_i + w_i) = \beta_0 + \beta_1 X_i + v_i$$  \hspace{1cm} (5)$$

This linear regression model is estimated by using OLS, ignoring the fact that $\tilde{R}_i$ is an imperfect measure of $\tilde{R}_i^T$, due to the fact that OLS is perfectly appropriate when the measurement error is uncorrelated with the explanatory variables. The OLS estimate with $\tilde{R}_i$ in place of $\tilde{R}_i^T$ produces consistent estimators of $\beta$ and the usual OLS inference procedures ($t$ statistics, $F$ statistics, etc.) are asymptotically valid. Assuming a single beta distribution over the business cycle, the conventional estimated prediction interval of $\tilde{R}_{t+1}$ given $X_{t+1}$ is:

$$IP_{\beta} = \left[ \tilde{R}_{t+1} \pm t_{n-k}^{\alpha/2} \left( \hat{\sigma}_v^2 \left( 1 + X_{t+1}'(XX)^{-1}X_{t+1} \right) \right)^{1/2} \right]$$  \hspace{1cm} (6)$$

where, $\hat{\sigma}_v^2$ is the OLS estimator of the error term variance and $t_{n-k}^{\alpha/2}$ is the $\alpha/2$ critical value of the $t$-Student distribution with $n-k$ degrees of freedom.

In the case where a mixture of two beta distributions is used to model the behavior of recovery rates over the business cycle, the corresponding estimated prediction interval of $\tilde{R}_{t+1}$ given $X_{t+1}$ is:

$$IP_{\text{mixture}} = \left[ \tilde{R}_{t+1} \pm t_{n-k}^{\alpha/2} \left( \hat{\sigma}_u^2 \left( 1 + X_{t+1}'(XX)^{-1}X_{t+1} \right) \right)^{1/2} \right]$$  \hspace{1cm} (7)$$

where, $\hat{\sigma}_u^2$ is the OLS estimator of the error term variance and $t_{n-k}^{\alpha/2}$ is the $\alpha/2$ critical value of the $t$-Student distribution with $n-k$ degrees of freedom.

It is expected that $IP_{\text{mixture}}$ improves the coverage accuracy of $IP_{\beta}$. Unfortunately, $IP_{\beta}$ and $IP_{\text{mixture}}$ are formed using the normal quantile transformation of the beta and the mixture distributed recovery values to convert them into a more normally distributed random variables, respectively. Nevertheless, as the error terms are
not normally distributed, both $IP_{\beta}$ and $IP_{\text{mixture}}$ could be very inaccurate and this inaccuracy might increase as sample size decreases (Lam and Veall, 2002).

One approach to improving the coverage accuracy of both $IP_{\beta}$ and $IP_{\text{mixture}}$ is to construct bootstrap prediction intervals (Efron and Tibshirani, 1993). The main steps for forming bootstrap prediction intervals may be summarized as follows. Under the assumption of beta distribution, generate $B$ bootstrap samples $\widetilde{R}_1^*, \ldots, \widetilde{R}_B^*$, where each $\widetilde{R}_b^*$, $b = 1, \ldots, B$, $t = 1, \ldots, n$, is calculated as $X' \hat{\beta}_{\text{OLS}}$ plus a bootstrap residual drawn randomly and with replacement from the OLS residuals of the original sample. Calculate new OLS estimate $\hat{\beta}^*$ and construct the bootstrap prediction error of each bootstrap sample as $v_{t+1}^* = \tilde{R}_{t+1}^* - (\hat{\beta}_0^* + \hat{\beta}_1^* X_{t+1})$. Let $v_{t+1}^* (\alpha/2)B$ -th order prediction error of the ordered bootstrap sample $v_{t+1}^* \leq \ldots \leq v_{t+1}^*$. Compute the bootstrap prediction interval as:

$$IP_{\beta}^* = [\widetilde{R}_{t+1}^* \pm v_{t+1}^* (\alpha/2)B]$$

Similarly, under the assumption of a mixture of two beta distributions:

$$IP_{\text{mixture}}^* = [\widetilde{R}_{t+1}^* \pm u_{t+1}^* (\alpha/2)B]$$

where, $u_{t+1}^* (\alpha/2)B$ is the $(\alpha/2)B$ -th order prediction error of the ordered bootstrap sample $u_{t+1}^* \leq \ldots \leq u_{t+1}^*$. As another candidate, the bootstrap percentile-t method can be used to construct prediction intervals for $\widetilde{R}_{t+1}^*$. This works by constructing the bootstrap version of $se_{\beta}$ and $se_{\text{mixture}}$, that is, $se_{\beta}^1, \ldots, se_{\beta}^B$ and $se_{\text{mixture}}^1, \ldots, se_{\text{mixture}}^B$, where $se_{\beta}^b = \hat{\sigma}_{\beta}$ and $se_{\text{mixture}}^b = \hat{\sigma}_{\text{mixture}}$, and replacing $t_{n-k}^{\alpha/2}$ in (6) and (7) with the $(\alpha/2)B$ -th order standarized prediction error of the ordered sample $e_{t+1}^* / se_{\beta}^1 \leq \ldots \leq e_{t+1}^* / se_{\beta}^B$, $(e_{t+1}^* / se_{\beta}^*)_{(\alpha/2)B}$, and with the $(\alpha/2)B$ -th order standarized prediction error of the
ordered sample \( e_{t+1}^*/se_{t}^{* mixture} \leq \ldots \leq e_{t+1}^*/se_{B}^{* mixture} \leq \ldots \leq e_{t+1}^*/se_{t}^{* mixture} \), respectively. The corresponding bootstrap percentile-t prediction intervals are given by:

\[
IP_{\text{beta}}^* = \left[ \bar{R}_{t+1} \pm (e_{t+1}^*/se_{t}^{*beta})_{(\alpha/2)B} \left( \tilde{\sigma}^2 \left[ 1 + X_{t+1}'(X'X)^{-1}X_{t+1} \right] \right)^{1/2} \right]
\]

(10)

\[
IP_{\text{mixture}}^* = \left[ \bar{R}_{t+1} \pm (e_{t+1}^*/se_{t}^{*mixture})_{(\alpha/2)B} \left( \tilde{\sigma}^2 \left[ 1 + X_{t+1}'(X'X)^{-1}X_{t+1} \right] \right)^{1/2} \right]
\]

(11)

III. Fitting finite mixtures of beta distributions

Most credit risk models assume that recovery rates are stochastic variables independent from default probabilities drawn from a beta distribution (for instance, PortfolioManager and CreditMetrics). There is no theoretical reason that this is the right shape, but it has been widely used to describe the observed behavior of the recovery rates because beta distribution is one of the few common “named” distributions that give probability 1 to a finite interval, here taken to be (0,1), corresponding to 100% loss or zero loss. Additionally, it has great flexibility in the sense that it is not restricted to being symmetrical. However, the pattern of the recovery rate distribution can vary significantly across seniority level and industries, and it may be questionable to use a beta distribution, which is calibrated on the mean and variance only (Hagmann et al., 2005). Also, the empirical distribution of the recovery rates may exhibit several local modes (Renault and Scaillet, 2004; Schüermann, 2004).

The beta distribution is indexed by two parameters and its probability density function is given by:

\[
f(y; p, q) = \frac{1}{B(p, q)} y^{p-1}(1 - y)^{q-1}, \quad 0 < y < 1, \quad p > 0, \quad q > 0
\]

(12)

where, \( B(\cdot, p, q) \) denotes the beta function, \( \alpha \) is the shape parameter and \( \beta \) is the scale parameter. Application of a beta distribution in LGD models can be found in Gordy and Jones (2002), Ivanova (2004), Onorota and Altman (2005) and Pesaran, et al. (2005), between others.
In this paper, we consider finite mixtures of beta distributions to model the observed behavior of LGD through a business cycle. The finite mixtures of beta distributions provide an extremely flexible method of modelling unknown distributional shapes over the business cycle, which apparently cannot be modelled by a single beta distribution function. With a mixture model-based approach, the recovery rates can be viewed as originating from \( g \)-component distributions in unknown mixing proportions \( \pi_j (j=1,\ldots,g) \), that are non-negative and sum to 1. The probability density function of the recovery rates is given by

\[
f(y; \pi, p, q) = \sum_{j=1}^{g} \pi_j f_j(y | p_j, q_j),
\]

in which \( \pi = (\pi_1, \ldots, \pi_{g-1}) \), \( p = (p_1, \ldots, p_g) \), \( q = (q_1, \ldots, q_g) \) and \( f_j(y | p_j, q_j), j=1,\ldots,g \), denotes the values of the univariate beta probability function specified by the parameters \( p_j \) and \( q_j \).

The mixing parameter \( \pi_j \) gives the prior probability that the security corresponds to the \( j \)th component of the mixture, representing an endogenous parameter which determines the relative importance of each component in the mixture of beta distributions. The posterior probability that security \( k \) comes from the component \( j \) is, from the Bayes rule,

\[
p_{kj} = \frac{\pi_j f_j(y_k | p_j, q_j)}{\sum_{j=1}^{g} \pi_j f_j(y_k | p_j, q_j)} = \frac{\pi_j f_j(y_k | p_j, q_j)}{f(y_k | p, q)},
\]

We use the called EM (Expectation-Maximization) algorithm (Dempster et al., 1977) to estimate the unknown parameters in the beta mixture model. Under the assumption that \( Y_1, \ldots, Y_n \) are independent and identically distributed random variables following a beta mixture distribution, the log-likelihood function is given by

\[
\log L(\pi, p, q | y) = \sum_{k=1}^{n} \log \left( \sum_{j=1}^{g} \pi_j f_j(y_k | p_j, q_j) \right),
\]
With the maximum likelihood approach to the estimation of \( \psi = (\pi, p, q) \), an estimate is provide by an appropriate root of the likelihood equation

\[
\frac{\partial \log L(\psi \mid y)}{\partial \psi} = 0
\]

(16)

The EM algorithm is used to find solutions of (16) corresponding to local maxima and it is guaranteed to converge to the MLE. Overall, it is based on the idea of replacing one difficult likelihood maximization with a sequence of easier maximizations whose limit is the answer to the original problem.

In the EM framework, the observed univariate data vector \( Y = (Y_1, \ldots, Y_n) \) is completed with a component-label vector \( Z = (Z_1, \ldots, Z_n) \). The label variable \( Z_{kj} = Z_k(j), \ k = 1, \ldots, n, \ j = 1, \ldots, g \), is 0 or 1 according to whether \( k \) corresponds to the business cycle \( j \). Hence, \( Z = (Z_1, \ldots, Z_n) \) is an unobservable vector of component-indicator variables, and \( Z_k, k = 1, \ldots, n \), are assumed to be independent random variables from a multinomial distribution consisting of one draw on \( g \) categories with respective probabilities \( \pi_1, \ldots, \pi_g \). That is,

\[
Z_1, \ldots, Z_n \sim \text{Mult}_g(1, \pi),
\]

(17)

where \( \pi = (\pi_1, \ldots, \pi_{g-1}) \). The complete-data log-likelihood is

\[
\log L_c(\pi, p, q \mid y) = \sum_{j=1}^{g} \left( \sum_{k=1}^{n} z_{kj} \log \{ \pi_j f_j(y_k; p_j, q_j) \} \right),
\]

(18)

The EM algorithm allows us to maximize \( L(\pi, p, q \mid y) \) by working with \( L_c(\pi, p, q \mid y, z) \). The EM algorithm is an iterative procedure. Each iteration comprises of the “E-step”, which calculates the expected log likelihood, and the “M-step”, which finds its maximum.

Now, we start the algorithm: From an initial value \( \psi^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) \), we create a sequence according to
\( \psi^{(r+1)} = \) the value that maximizes \( E[\log L(\psi | y, z) | \psi^{(r)}, y] = Q(\psi; \psi^{(r)}) \), \( (19) \)

the conditional expectation of the complete data log-likelihood \( \log L_c(\pi, p, q | y, z) \), given the observed data \( y \), using the current fit \( \psi^{(r)} \) for \( \psi \).

On the \((r+1)\) iteration, the E-step requires the calculation of \( Q(\psi; \psi^{(r)}) \). Since \( Z = (Z_1, ..., Z_n) \) is non observed data, the E-step is effected by replacing \( z_{kj} \) by its conditional expectation given \( y_j \), using \( \psi^{(r)} \) for \( \psi \). That is, \( z_{kj} \) is replaced by \( \tau_j(y_k : \psi^{(r)}) = E_{\psi^{(r)}}[Z_{kj} | y_k] = \Pr_{\psi^{(r)}}[Z_{kj} = 1 | y_k] = p_{kj} \). On the M-step, on the \((r+1)\) iteration we choose the value of \( \psi \), say \( \psi^{(r+1)} \), that maximizes \( Q(\psi; \psi^{(r)}) \). Then, the vector \( \psi^{(r+1)} \) is obtained as an appropriate root of

\[
\sum_{j=1}^{g} \sum_{k=1}^{n} \tau_j(y_k, \psi^{(r)}) \frac{\partial \log L(\psi_j | y)}{\partial \psi} = 0 \quad (20)
\]

The hidden key to the algorithm is the application of the information inequality (Demspter et al., 1977, lemma 1), which states that \( L(\hat{\psi}^{(r+1)} | y) \geq L(\hat{\psi}^{(r)} | y) \), with equality holding if and only if successive iterations yield the same value of the maximized expected complete-data log-likelihood, that is, \( E[\log L(\psi^{(r+1)} | y, z) | \psi^{(r)}, y] = E[\log L(\psi^{(r)} | y, z) | \psi^{(r)}, y] \).

On consequence, the E-step and the M-step are alternated repeatedly until the likelihood changes by an arbitrarily small amount in the case of convergence.

An important issue involves the choice of the number \( g \) of beta component densities. There is not a priori information regarding the number of components in the beta mixture model. The selection of the number \( g \) of component densities compatible with the data should follow a statistical criterion. The most natural way is to test the null hypothesis \( H_0: g = g_0 \) versus \( H_1: g = g_1 \), for \( g_1 > g_0 \) by the likelihood ratio statistic (LR)

\[
-2 \log \lambda = 2[\log L(\hat{\psi}_1) - \log L(\hat{\psi}_0)]. \quad (21)
\]
However, the standard LR is not asymptotically chi-squared distributed, as $H_0$ corresponds to a non-identifiable subset of the parameter space (McLachlan, 1987, Lo et al., 2001). Hence, we use the following bootstrap procedure to estimate the null distribution of LR:

**Step 1:** Compute the original test statistic $\hat{T} = -2 \log \lambda$ using the original sample.

**Step 2:** Under the null hypothesis of $g_0$ beta components, $B$ bootstrap samples are generated parametrically. For each bootstrap sample, the beta mixture model is fitted for both $g=g_0$ and $g=g_1$, and the bootstrap version of the LR is evaluated. In this way, a sample of $B$ independent (conditional on the original sample) observations of $\hat{T}$, say $\hat{T}_1^*, \ldots, \hat{T}_B^*$, is obtained.

**Step 3:** Let $\hat{T}_{(1-\alpha)B}^*$ the $(1-\alpha)B$-th order statistic of the sample $\hat{T}_1^*, \ldots, \hat{T}_B^*$, given a significance level $\alpha$. Reject the null hypothesis at the significance level $\alpha$ if $\hat{T} > \hat{T}_{(1-\alpha)B}^*$.

**Step 4:** Compute the bootstrap $p$-value as $p_B = \text{card}(\hat{T}_b^* \geq T^*) / B$, $b = 1, \ldots, B$.

**IV. Simulation study**

We perform Monte Carlo experiments to shed some light on the performance of prediction intervals under the assumption of the beta distribution over the business cycle. Alternative prediction intervals are provided assuming mixtures of two beta distributions, distinguishing between secured and unsecured securities which are strongly or weakly sensitive to the business cycle. Furthermore, percentile bootstrap and percentile-t bootstrap prediction intervals are constructed.

The simulation experiment works as follows. We generate 1000 samples of $R_t$, $t=1,\ldots,1000+1$ from a mixture of two beta distributions. The simulations are based on unsecured and secured securities passing from an expansion period to a recession period, considering two possibilities: the unsecured/secured security is strongly sensitive to the business cycle and it is weakly sensitive to the business cycle.

Overall:
1. Unsecured security (Expansion period to recession period):
   1.1. Strongly sensitive to the business cycle.
   
   \[
   f_{\text{Un}}^S(y | \pi_E, \pi_R, p_E, q_E, p_R, q_R) = \pi_E f_E(y | p_E, q_E) + \pi_R f_R(y | p_R, q_R)
   \]  
   (22)

   1.2. Weakly sensitive to the business cycle.
   
   \[
   f_{\text{Un}}^W(y | \pi_E, \pi_R, p_E, q_E, p_R, q_R) = \pi_E f_E(y | p_E, q_E) + \pi_R f_R(y | p_R, q_R)
   \]  
   (23)

2. Secured security (Expansion period to recession period):
   2.1. Strongly sensitive to the business cycle.
   
   \[
   f_{\text{Sc}}^S(y | \pi_E, \pi_R, p_E, q_E, p_R, q_R) = \pi_E f_E(y | p_E, q_E) + \pi_R f_R(y | p_R, q_R)
   \]  
   (24)

   2.2. Weakly sensitive to the business cycle.
   
   \[
   f_{\text{Sc}}^W(y | \pi_E, \pi_R, p_E, q_E, p_R, q_R) = \pi_E f_E(y | p_E, q_E) + \pi_R f_R(y | p_R, q_R)
   \]  
   (25)

In our analysis we consider that \( \pi_E = \pi_R = 0.5 \), which means that expansion periods and recession periods have the same probability. Figure 1 shows the four mixtures of beta distributions we use in the Monte Carlo design.
We generate data which represent strongly sensitive to the business cycle unsecured securities from 
$0.5f_E(y | p_E = 1.8, q_E = 8) + 0.5f_R(y | p_R = 5, q_R = 5)$, where $f_E(y | p_E = 1.8, q_E = 8)$ and $f_R(y | p_R = 5, q_R = 5)$ represent the observed behaviour of those securities conditioned to recession and expansion periods, respectively. In the case where the unsecured securities are weakly sensitive to the business cycle, the data are drawn from the mixture $0.5f_E(y | p_E = 1.8, q_E = 8) + 0.5f_R(y | p_R = 4, q_R = 7)$, where $f_E(y | p_E = 4, q_E = 7)$ shows the shape of the distribution of those securities in...
expansion period. On the other hand, we use the mixture of beta distributions 
\[ 0.5 f_E(y | p_E = 8, q_E = 1.8) + 0.5 f_R(y | p_R = 5, q_R = 5) \] to generate data which represent strongly sensitive to the business cycle secured securities. Finally, the mixture 
\[ 0.5 f_E(y | p_E = 8, q_E = 1.8) + 0.5 f_R(y | p_R = 7, q_R = 4) \] is used to draw secured securities which are weakly sensitive to the business cycle, as it can be observed in Figure 1.

Table 1 reports the mean of the beta distributions we use as components of the mixtures of beta distributions presented in Figure 1. These values refer to the conditional mean of secured/unsecured securities (conditioned to an expansion or recession period), classified by their sensitivity to the business cycle. The unconditional mean (unconditioned to the business cycle) is also reported in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Expansion</th>
<th>Recession</th>
<th>Mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unsecured</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly sensitive</td>
<td>50%</td>
<td>18.3%</td>
<td>34.1%</td>
</tr>
<tr>
<td>Weakly sensitive</td>
<td>36.3%</td>
<td>18.3%</td>
<td>27.3%</td>
</tr>
<tr>
<td><strong>Secured security</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly sensitive</td>
<td>81.6%</td>
<td>50%</td>
<td>65.8%</td>
</tr>
<tr>
<td>Weakly sensitive</td>
<td>81.6%</td>
<td>63.6%</td>
<td>72.6%</td>
</tr>
</tbody>
</table>

Under each scenario, firstly we apply the MKMV approach and construct \( IP_{\text{beta}} \). Secondly, we assume that the data follows a mixture of two beta distributions, estimate the unknown parameters \( \pi, p_E, q_E, p_R, q_R \) by EM algorithm and construct \( IP_{\text{mixture}} \). Lastly, we design bootstrap versions of both, \( IP_{\text{beta}} \) and \( IP_{\text{mixture}} \) based on \( B=1000 \) bootstrap samples, using the percentile bootstrap and the percentile-t bootstrap. In this way, we construct \( IP_{\text{beta}}^*, IP_{\text{mixture}}^* \) (percentile bootstrap prediction intervals) and \( IP_{\text{beta}}^{t*}, IP_{\text{mixture}}^{t*} \) (percentile-t bootstrap prediction intervals).

We treat the six prediction intervals, \( IP_{\text{beta}}, IP_{\text{mixture}}, IP_{\text{beta}}^*, IP_{\text{mixture}}^*, IP_{\text{beta}}^{t*}, IP_{\text{mixture}}^{t*} \) at different significant levels, from the 90% to the 99% significant level, to compare their performance when the significant level varies. We compute the percentage of forecasts of \( R_{1000}^T \) conditional upon \( X_{1001} \) that are within the prediction intervals over the 1000 trials.
Finally, we conduct the evaluation of prediction interval accuracy. For any prediction interval, the coverage error is defined as the difference between the proportion of failures $\hat{\alpha}$ (times that the forecast lies out of the prediction interval) and the nominal significance level $\alpha$. In order to check if the difference between $\hat{\alpha}$ and $\alpha$ is statistically significant, we use a method based on the binomial distribution (Kupiec, 1995). The method proposed by Kupiec (1995) is based on the assumption that the estimated prediction interval is accurate, so the observations of the events “forecast lies out of the prediction interval” and “forecast does not lie out of the prediction interval” can be modelled as draws from an independent binomial random variable with probability of occurrence equal to a specified $\alpha$ percent. The null hypothesis “the empirical size of the test is equal to the nominal size” is tested using the following likelihood ratio test statistic based on the binomial distribution:

$$LR_{PF} = 2\left[\ln\left(\hat{\alpha}^z (1 - \hat{\alpha})^{T-z}\right) - \ln\left(\alpha^z (1 - \alpha)^{T-z}\right)\right], \quad (26)$$

where, $z$ denotes the number of times the forecast lies out of the prediction interval and $\hat{\alpha}$ is the estimated empirical size. Under the null hypothesis, $LR_{PF}$ is asymptotically distributed as $\chi^2(1)$.

V. Numerical results

We present the results from a simulation study comparing performances of the six prediction intervals considered here. Specifically, we compare the coverage probabilities of the prediction intervals, $IP_{\text{beta}}$, $IP_{\text{mixture}}$, $IP^{*}_{\text{beta}}$, $IP^{*}_{\text{mixture}}$, $IP^{*}_{\text{beta}}$ and $IP^{*}_{\text{mixture}}$, for a nominal level varying from $\alpha = 0.01$ to $\alpha = 0.1$, in increments of 0.01.

Table 2 provides coverage properties of the competing prediction intervals to predict $R_{t+1}$ for unsecured securities which are strongly sensitive to the business cycle.
Table 2. Prediction Intervals for strongly sensitive unsecured securities

<table>
<thead>
<tr>
<th>α</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\alpha} - \alpha)</th>
<th>p-value (LR_{PF})</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\alpha} - \alpha)</th>
<th>p-value (LR_{PF})</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\alpha} - \alpha)</th>
<th>p-value (LR_{PF})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.050</td>
<td>0.040</td>
<td>0.0000</td>
<td>0.037</td>
<td>-0.027</td>
<td>0.0000</td>
<td>0.034</td>
<td>0.024</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.02</td>
<td>0.054</td>
<td>0.034</td>
<td>0.0000</td>
<td>0.062</td>
<td>-0.042</td>
<td>0.0000</td>
<td>0.060</td>
<td>0.040</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.03</td>
<td>0.059</td>
<td>0.029</td>
<td>0.0000</td>
<td>0.079</td>
<td>-0.049</td>
<td>0.0000</td>
<td>0.079</td>
<td>0.049</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.04</td>
<td>0.064</td>
<td>0.024</td>
<td>0.0003</td>
<td>0.102</td>
<td>-0.062</td>
<td>0.0000</td>
<td>0.101</td>
<td>0.061</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.064</td>
<td>0.014</td>
<td>0.0510</td>
<td>0.123</td>
<td>-0.073</td>
<td>0.0000</td>
<td>0.117</td>
<td>0.067</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.06</td>
<td>0.066</td>
<td>0.006</td>
<td>0.4312</td>
<td>0.152</td>
<td>-0.092</td>
<td>0.0000</td>
<td>0.149</td>
<td>0.089</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.07</td>
<td>0.073</td>
<td>0.003</td>
<td>0.7118</td>
<td>0.197</td>
<td>-0.127</td>
<td>0.0000</td>
<td>0.182</td>
<td>0.112</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.08</td>
<td>0.076</td>
<td>-0.004</td>
<td>0.6384</td>
<td>0.251</td>
<td>-0.171</td>
<td>0.0000</td>
<td>0.228</td>
<td>0.148</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.09</td>
<td>0.080</td>
<td>-0.010</td>
<td>0.2608</td>
<td>0.322</td>
<td>-0.232</td>
<td>0.0000</td>
<td>0.286</td>
<td>0.196</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.10</td>
<td>0.080</td>
<td>-0.020</td>
<td>0.2924</td>
<td>0.365</td>
<td>-0.265</td>
<td>0.0000</td>
<td>0.336</td>
<td>0.236</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 2 consists of blocks where each line provides the percentage of times \(R_{t+1}\) lies out the prediction interval, \(\hat{\alpha}\), the coverage error, defined as \(\hat{\alpha} - \alpha\), as well as the p-value of the \(LR_{PF}\) test for each of the six prediction intervals considered here. It can be observed that \(IP_{\beta}\) does not present good coverage properties for significance levels vary from 0.01 to 0.05. Using \(IP_{mixture}\) provides coverage well below the nominal level in most cases. The bad coverage characteristics of both \(IP_{\beta}\) and \(IP_{mixture}\) are due to the fact that the assumption of normality is not held. Note that \(IP_{mixture}^*\) and \(IP_{mixture}^*\) improve the coverage properties of the previous prediction intervals in contrast to \(IP_{\beta}^*\) and \(IP_{\beta}^*\). Both \(IP_{mixture}^*\) and \(IP_{mixture}^*\) exhibits the best coverage characteristics with respect to approximation to the nominal level.

Table 3 reports the results we obtain for the case of weakly sensitive to the business cycle unsecured securities.
Table 3. Prediction Intervals for weakly sensitive unsecured securities

<table>
<thead>
<tr>
<th>α</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\alpha} - \alpha)</th>
<th>p-value LRPF</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\alpha} - \alpha)</th>
<th>p-value LRPF</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\alpha} - \alpha)</th>
<th>p-value LRPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.021</td>
<td>-0.011</td>
<td>0.0023</td>
<td>0.015</td>
<td>0.005</td>
<td>0.1389</td>
<td>0.014</td>
<td>0.004</td>
<td>0.2305</td>
</tr>
<tr>
<td>0.02</td>
<td>0.027</td>
<td>0.007</td>
<td>0.1331</td>
<td>0.036</td>
<td>0.016</td>
<td>0.0011</td>
<td>0.038</td>
<td>0.018</td>
<td>0.0002</td>
</tr>
<tr>
<td>0.03</td>
<td>0.030</td>
<td>-0.000</td>
<td>1.0000</td>
<td>0.047</td>
<td>0.017</td>
<td>0.0035</td>
<td>0.047</td>
<td>0.017</td>
<td>0.0035</td>
</tr>
<tr>
<td>0.04</td>
<td>0.033</td>
<td>-0.007</td>
<td>0.2445</td>
<td>0.061</td>
<td>0.021</td>
<td>0.0016</td>
<td>0.061</td>
<td>0.021</td>
<td>0.0016</td>
</tr>
<tr>
<td>0.05</td>
<td>0.036</td>
<td>-0.014</td>
<td>0.0328</td>
<td>0.075</td>
<td>0.025</td>
<td>0.0007</td>
<td>0.074</td>
<td>0.024</td>
<td>0.0011</td>
</tr>
<tr>
<td>0.06</td>
<td>0.041</td>
<td>-0.019</td>
<td>0.0074</td>
<td>0.094</td>
<td>0.034</td>
<td>0.0000</td>
<td>0.087</td>
<td>0.027</td>
<td>0.0007</td>
</tr>
<tr>
<td>0.07</td>
<td>0.042</td>
<td>-0.028</td>
<td>0.0001</td>
<td>0.105</td>
<td>0.035</td>
<td>0.0000</td>
<td>0.099</td>
<td>0.029</td>
<td>0.0006</td>
</tr>
<tr>
<td>0.08</td>
<td>0.046</td>
<td>-0.034</td>
<td>0.0000</td>
<td>0.120</td>
<td>0.040</td>
<td>0.0000</td>
<td>0.110</td>
<td>0.030</td>
<td>0.0008</td>
</tr>
<tr>
<td>0.09</td>
<td>0.050</td>
<td>-0.040</td>
<td>0.0000</td>
<td>0.141</td>
<td>0.051</td>
<td>0.0000</td>
<td>0.122</td>
<td>0.032</td>
<td>0.0007</td>
</tr>
<tr>
<td>0.10</td>
<td>0.055</td>
<td>-0.045</td>
<td>0.0000</td>
<td>0.173</td>
<td>0.073</td>
<td>0.0000</td>
<td>0.138</td>
<td>0.038</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 3 shows that \(IP_{\beta}\) and \(IP_{\text{mixture}}\) do not yield acceptable coverage properties. In both cases, the percentage of rejections \(\hat{\alpha}\) tends to be below \(\alpha\) as \(\alpha\) increases. As it can be observed, their coverage properties are widely improved by using \(IP^*_{\text{mixture}}\) and \(IP^*_{\text{mixture}}\).

The same analysis is carried out in the case where the securities are secured. Tables 4 and 5 summarize the results corresponding to strongly sensitive to the business cycle secured securities and weakly sensitive to the business cycle secured securities, respectively.
### Table 4. Prediction Intervals for strongly sensitive secured securities

<table>
<thead>
<tr>
<th>α</th>
<th>IP_{\beta}</th>
<th>p-value</th>
<th>IP_{\beta}^{*}</th>
<th>p-value</th>
<th>IP_{\beta}^{**}</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>ˆα</td>
<td>ˆα – α</td>
<td>LR_{PF}</td>
<td>ˆα</td>
<td>ˆα – α</td>
</tr>
<tr>
<td>0.01</td>
<td>0.050</td>
<td>0.040</td>
<td>0.0000</td>
<td>0.035</td>
<td>0.025</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.02</td>
<td>0.057</td>
<td>0.037</td>
<td>0.0000</td>
<td>0.059</td>
<td>0.039</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.03</td>
<td>0.063</td>
<td>0.033</td>
<td>0.0000</td>
<td>0.075</td>
<td>0.045</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.04</td>
<td>0.066</td>
<td>0.026</td>
<td>0.0001</td>
<td>0.097</td>
<td>0.057</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.069</td>
<td>0.019</td>
<td>0.0089</td>
<td>0.123</td>
<td>0.073</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.06</td>
<td>0.073</td>
<td>0.013</td>
<td>0.0934</td>
<td>0.156</td>
<td>0.096</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.07</td>
<td>0.076</td>
<td>0.006</td>
<td>0.4628</td>
<td>0.197</td>
<td>0.127</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.08</td>
<td>0.076</td>
<td>-0.004</td>
<td>0.6384</td>
<td>0.245</td>
<td>0.165</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.09</td>
<td>0.076</td>
<td>-0.014</td>
<td>0.1127</td>
<td>0.307</td>
<td>0.217</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.10</td>
<td>0.078</td>
<td>-0.022</td>
<td>0.0162</td>
<td>0.363</td>
<td>0.263</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α</th>
<th>IP_{\text{mixture}}</th>
<th>p-value</th>
<th>IP_{\text{mixture}}^{*}</th>
<th>p-value</th>
<th>IP_{\text{mixture}}^{**}</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>ˆα</td>
<td>ˆα – α</td>
<td>LR_{PF}</td>
<td>ˆα</td>
<td>ˆα – α</td>
</tr>
<tr>
<td>0.01</td>
<td>0.029</td>
<td>0.019</td>
<td>0.0000</td>
<td>0.006</td>
<td>-0.004</td>
<td>0.1696</td>
</tr>
<tr>
<td>0.02</td>
<td>0.033</td>
<td>0.013</td>
<td>0.0071</td>
<td>0.019</td>
<td>-0.001</td>
<td>0.8198</td>
</tr>
<tr>
<td>0.03</td>
<td>0.037</td>
<td>0.007</td>
<td>0.2102</td>
<td>0.032</td>
<td>0.002</td>
<td>0.7137</td>
</tr>
<tr>
<td>0.04</td>
<td>0.041</td>
<td>0.001</td>
<td>0.8723</td>
<td>0.042</td>
<td>0.002</td>
<td>0.7487</td>
</tr>
<tr>
<td>0.05</td>
<td>0.046</td>
<td>-0.004</td>
<td>0.5565</td>
<td>0.054</td>
<td>0.004</td>
<td>0.5664</td>
</tr>
<tr>
<td>0.06</td>
<td>0.047</td>
<td>-0.013</td>
<td>0.0725</td>
<td>0.067</td>
<td>0.007</td>
<td>0.3597</td>
</tr>
<tr>
<td>0.07</td>
<td>0.048</td>
<td>-0.022</td>
<td>0.0039</td>
<td>0.075</td>
<td>0.005</td>
<td>0.5398</td>
</tr>
<tr>
<td>0.08</td>
<td>0.050</td>
<td>-0.030</td>
<td>0.0001</td>
<td>0.085</td>
<td>0.005</td>
<td>0.5636</td>
</tr>
<tr>
<td>0.09</td>
<td>0.052</td>
<td>-0.038</td>
<td>0.0000</td>
<td>0.089</td>
<td>-0.001</td>
<td>0.9118</td>
</tr>
<tr>
<td>0.10</td>
<td>0.055</td>
<td>-0.045</td>
<td>0.0000</td>
<td>0.092</td>
<td>-0.008</td>
<td>0.3933</td>
</tr>
</tbody>
</table>
Owing to the underlying error terms are not normally distributed, \( IP_{\text{beta}} \) and \( IP_{\text{mixture}} \) do not perform well, even though slightly better results are obtained choosing \( IP_{\text{mixture}} \). As one would expect, the bootstrap versions \( IP^*_{\text{mixture}} \) and \( IP^{**}_{\text{mixture}} \) are the most accurate in terms of approximation of \( \hat{\alpha} \) to \( \alpha \). The null hypothesis “the empirical size of the test is equal to the nominal size” is accepted at 5% significance level in almost all cases as the p-value of the \( LR_{\text{PF}} \) test indicates.

To sum up the simulation results indicate that both \( IP^*_{\text{mixture}} \) and \( IP^{**}_{\text{mixture}} \) yield percentages of rejections most consistently close to \( \alpha \) for all the different values of \( \alpha \) considered here. The results presented in Tables 2-5 are indicative of the types of results that may generally be obtained upon the systematic use of a single beta distribution over the business cycle to make prediction intervals.
Finally, in order to check the behaviour of the likelihood ratio test statistic (LR) to test the number of the component beta densities we perform a Monte Carlo experiment. In such experiment we test the null hypothesis that a random sample is drawn form a beta distribution, \( g=1 \), versus the alternative hypothesis is that the sample is drawn from a mixture of two beta distributions, \( g=2 \). We use two approaches. The first one is to test \( H_0: g=1 \) versus \( H_1: g=2 \) using the likelihood ratio test with an asymptotic chi-squared distribution. The second one is to use the bootstrap distribution of the likelihood ratio test, based on \( B=500 \) bootstrap replications.

In Table 6 we report the proportion of rejections of \( H_0 \) at 1%, 5% and 10% significance levels in 1000 simulation runs. This table provides empirical size and power properties of the two competing approaches to test the hypotheses \( H_0: g=1 \) versus \( H_1: g=2 \). Sample size \( n=1000 \) is considered.

Table 6. Results of testing the number of \( g \)-component beta densities

<table>
<thead>
<tr>
<th>Data generating process</th>
<th>( \alpha=1% )</th>
<th>( \alpha=5% )</th>
<th>( \alpha=10% )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \chi^2 )-test</td>
<td>Bootstrap</td>
<td>( \chi^2 )-test</td>
</tr>
<tr>
<td>( b(8,1.8) )</td>
<td>28.2*</td>
<td>1.8</td>
<td>29.9*</td>
</tr>
<tr>
<td>( b(7,4) )</td>
<td>30.6*</td>
<td>0.6</td>
<td>31.5*</td>
</tr>
<tr>
<td>( b(5,5) )</td>
<td>27.1*</td>
<td>0.6</td>
<td>27.6*</td>
</tr>
<tr>
<td>( 0.5b(1.8,8) + 0.5b(5,5) )</td>
<td>99.3</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td>( 0.5b(8.1,8) + 0.5b(7,4) )</td>
<td>34.5</td>
<td>58.6</td>
<td>59.4</td>
</tr>
<tr>
<td>( 0.5b(8.1,8) + 0.5b(5,5) )</td>
<td>99.3</td>
<td>99.8</td>
<td>99.9</td>
</tr>
<tr>
<td>( 0.5b(8,1.8) + 0.5b(4,7) )</td>
<td>35.2</td>
<td>58.6</td>
<td>59</td>
</tr>
</tbody>
</table>

* significant at 1% level.

Table 6 consists of two blocks where each line provides the proportion of rejection of \( H_0 \) obtained from each test statistic. Similarly, columnwise entries correspond to particular data generating processes employed for the simulation. The empirical size of the \( \chi^2 \)-test and the bootstrap test are shown in the first block of Table 6. It can be observed that the \( \chi^2 \)-test rejects the null hypothesis much more often than does the bootstrap test at all significance levels considered here. To diagnose the discrepancies between the empirical size and the significance level of the test which are
significant at the 1% level we use the significance test statistic \( Z = \frac{(p - \alpha)}{\sqrt{\alpha(1 - \alpha)/R}} \), where \( p \) is a rejection rate, \( \alpha \) is the significance level and \( R \) is the number of replications. Hence, significant size distortions are diagnosed in case the empirical rejection frequencies obtained under the null hypothesis are not covered by a confidence band constructed around the significant level as \( \alpha \pm 2.575 \frac{(p - \alpha)}{\sqrt{\alpha(1 - \alpha)/R}} \). In our simulations, a test statistic is well specified if \( p \in [0.1898,1.8102] \), \( p \in [3.226,6.774] \) and \( p \in [7.558,12.442] \) for \( R=1000 \), at \( \alpha = 1\% \), \( \alpha = 5\% \) and \( \alpha = 10\% \), respectively. It is worthwhile to point out that the bootstrap-based test statistic yields a correct empirical size in all cases while the \( \chi^2 \)-based test statistic always shows significant oversizing.

The observed power of the tests is estimated from 1000 samples drawn from a two-component beta mixture alternative, based on a mixing proportion \( \pi = 0.5 \). The bootstrap-based test statistic behaves extremely well in terms of power. The results also indicate that the \( \chi^2 \)-based test statistic is quite powerful too.

The results in Table 6 allow us to give a clear-cut answer to the question of what test should be preferred to test the number \( g \) of components in a beta mixture model over the business cycle. Bootstrap-based test statistic performs absolutely well in terms of size and power in our experiment. On the other hand, the \( \chi^2 \)-based test statistic behaves quite well in terms of power, but the asymptotic critical value is not accurate at all, which results in a large discrepancy between empirical size and significance level of the test.

VI. Conclusions

Our simulations indicate that, under the Moody’s KMV model context, the prediction intervals based on the widespread assumption of beta distribution over the business cycle do not perform well at all. Alternative prediction intervals constructed assuming mixtures of two beta distributions over the business cycle do not improve the accuracy of beta prediction intervals. The underlying reason is the fact that the error terms of the regression model are not normally distributed. However, the coverage
accuracy of $IP_{\text{beta}}$ and $IP_{\text{mixture}}$ is exceptionally improved by constructing bootstrap versions of $IP_{\text{mixture}}$ in contrast to happen if bootstrap version of $IP_{\text{beta}}$ are formed.

A relevant conclusion that follows from our experiments is that loss given default models should select mixtures of beta distributions to model the observed behaviour of secured/unsecured securities over the business cycle since apparently their empirical distribution over the business cycle cannot be modelled by a single beta distribution. Moreover, a bootstrap-based test statistic should be employed to test the number $g$ of components in a beta mixture model over the business cycle. Another relevant conclusion is that bootstrap versions of mixture prediction intervals should be constructed in order to achieve accurate LGD prediction intervals. The results of our experiments suggest that it is possible to consider either $IP_{\text{mixture}}^s$ or $IP_{\text{mixture}}^p$ to improve the accuracy of beta prediction intervals, since in practice we cannot know in advance which of them would lead to a more accurate prediction interval. To sum up, the standard assumption of beta distribution to model the observed behaviour of secured/unsecured securities over the business cycle should therefore be considered with caution on the calculus of LGD prediction intervals.
References


