Tech Science Press

Estimating Loss Given Default Based on Beta Regression

**Jamil J. Jaber[1,2], Noriszura Ismail[2], Siti Norafidah Mohd Ramli[2],
Baker Albadareen[2] and Nawaf N. Hamadneh[3,*]**

[1]Department of Risk Management and Insurance, The University of Jordan, Aqaba, 77111, Jordan
[2]Universiti Kebangsaan Malaysia, Department of Mathematical Sciences, Malaysia
[3]Department of Basic Sciences, College of Science and Theoretical Studies, Saudi Electronic University,
Riyadh, 11673, Saudi Arabia
*Corresponding Author: Nawaf N. Hamadneh. Email: nwwaf977@gmail.com

Abstract: Loss given default (LGD) is a key parameter in credit risk management to calculate the required regulatory minimum capital. The internal ratings-based (IRB) approach under the Basel II allows institutions to determine the loss given default (LGD) on their own. In this study, we have estimated LGD for a credit portfolio data by using beta regression with precision parameter ($\varnothing$) and mean parameter (μ). The credit portfolio data was obtained from a banking institution in Jordan; for the period of January 2010 until December 2014. In the first stage, we have used the "outstanding amount" and "amount of borrowing" to find LGD of each default borrower (494 out of 4393 borrower). In the second stage, we fit univariate parametric distributions to the LGD data to obtain the beta distribution. After that, we have estimated the values of $\varnothing$ based on microeconomic variables (SPP, OE and LR). Moreover, we have estimated the values of μ based on macroeconomic variables (GDP and Inflation rate). Finally, we have compared between six different link functions (Logit, loglog, probit, cloglog, cauchit, and log), which have used with $\varnothing$ and μ. The results show that Beta regression with probit link function has the highest R-squared with accepted measurements for logL, AIC and BIC.

Keywords: Credit risk; loss given default; parametric distribution; regression model

1 Introduction

The Basel Committee gives three approaches to estimate the required regulatory capital in banking institutions; standardized, internal ratings-based (IRB) and progressing IRB. The IRB approach is generally favored compared to the standard approach because it produces higher accuracy estimates and lower capital charges. The IRB approach under the Basel II allows institutions to determine probabilities of default (PD) and loss given default (LGD) on their own, as opposed to the standardized approach under the Basel I which uses estimates of PD and LGD from the Central Bank to calculate the required capital based on a percentage of risk-weighted-assets. Thus,

the Basel II leads to a better differentiation of risks and considers diversification of a bank's portfolio [1,2].

In Jordan, the banking sector obtains the estimate of LGD from the Central Bank of Jordan under the Basel I. Therefore, the LGD estimate is fixed, and does not varies according to banks. The contributions of this chapter are to estimate the LGD under the Basel II for corporate credit portfolio in the Jordanian banking sector, to fit the LGD data to parametric distributions, and to incorporate internal and external financial variables into the data so that the LGD data can be fitted to regression models. The main advantage of fitting the LGD data with financial variables as covariates is that we can determine which financial variables significantly affect the LGD. Since the LGD is influenced by some key transaction characteristics, several categories of variables such as macroeconomic and industry-specific variables can be used to build a predictive regression model.

The objective of this chapter is to use several parametric models (beta, Cauchy, gamma, Gompertz, logistic, log-normal, gamma, normal, and Weibull) to estimate LGD. The models are fitted to a sample data obtained from the corporate credit portfolio of a bank in Jordan for the period of January 2010 until December 2014. The LGD data lies between interval [0,1] because it is the proportion of outstanding amount from the borrowing amount. We also consider five financial variables for the regression model to determine the financial variables, which significantly affect the LGD. The financial variables are gross domestic product (GDP), inflation rate (Inf), service pricing policy (SPP), operating efficiency (OE), and liquidity ratio (LR).

The rest of this paper proceeds as follows. In the next section, we discuss the literature review for important work for LGD, and Section 3 describes the parametric distributions for fitting the LGD data and the regression models for determining the financial variables which significantly affect the LGD. We present the sample data and the empirical results in Section 4, and the final section concludes.

2 Literature Review

Most empirical studies on credit risk depended heavily on corporate bond markets to gauge losses in the case of default. The reason behind this is that bank loans are private instruments, and thus, little information on loan losses are freely accessible. Several studies on credit risks and LGD on bond markets have been carried out in the last several decades. An earlier study can be found in [3] who utilized actuarial analysis to investigate mortality rates of U.S. corporate bonds. This was followed by various empirical studies on credit risk in bond markets (see, for example [4,5]). The mortality approach was also used by [6] to measure the percentage of bad and doubtful loans of corporate bond recovered several months after the default date. Recent studies on LGD can be found in [7] who proposed a new model for LGD of bank loans by leveraging time to recovery, and reference [8] who forecasted LGD of bank loans using multi-stage model. In another study, [9] constructed a survival model to predict risks of cardholders and applied the model to a case study in Capital Card Services.

LGD data from banking industry tends to be skewed and heavy-tailed, and thus, can be fitted by parametric models such as beta, lognormal, gamma, and Pareto. Besides parametric models, non-parametric models were also proposed and used, such as regression tree, neural networks, multivariate adaptive regression spline, and least squares support vector machine [10–17].

For the case of LGD with covariates, regression models can be utilized and several examples can be found from past studies. Examples of regression models for LGD data are ordinary least squares regression (OLS), ridge regression (RiR), and fractional response regression. In 2005 Moody's introduced the renowned LossCals Model using a multivariate linear regression model consisting of industry and macroeconomic factors, and reference [18] applied logistic regression with time consideration using transformed LGD as dependent variable and macroeconomic variables as independent variables. Recently, reference [19] used quintiles regression to estimate downturn and unexpected credit losses known as downturn LGD. Finally, reference [9] constructed a model to predict the risk of a cardholder for the lifetime of account and applied survival analysis methodologies to a case study in capital card services.

Under the Basel II, banking institutions are suggested to consider macroeconomic downturn conditions when estimating recovery rates [20–22]. In particular, reference [22] assumes that banking institutions should use gross domestic product (GDP) growth rate and unemployment rate as determinants of recovery rate prediction. It should be noted that recovery rate is equal to one minus LGD rate. Studies from [23,24] showed that GDP growth rate was significantly relevant to the recovery rate of the U.S. bonds. On the other hand, references [12,25] found that GDP growth rate was not significantly relevant.

Other macroeconomic covariates were also suggested in the literature to predict recovery rate, such as inflation rate [24], interest rate [14,24], growth rate of investment [12,26] and rate of return on stock market [24,25]. Studies from [6,27] showed that recovery rate decreases when loan size increases. In another study, reference [8] used Japanese credit portfolio to analyze impacting factors of LGD and to improve multi-stage model for predicting LGD. The variables considered are creditworthiness score, collateral quota (commercial bills), collateral quota (real estate), collateral quota (marketable securities), collateral quota (deposits), credit guarantee quota, and exposure (in hundred million yen). Their results showed that collateral, guarantees, and loan size significantly affect the LGD.

In this study, we consider macro- and micro-economics factors as explanatory variables, and the factors are obtained from available reports. We use macro-economic factors (GDP and inflation rate) for the mean parameter and micro-economic factors (service pricing policy, operating efficiency, and liquidity ratio) for the dispersion parameter in beta regression model.

3 Methodology for Estimating LGD

This section gives a background of the main concepts used in our study. Our model consists of three stages. In the first stage, we use the outstanding amount and amount of borrowing to find LGD of each default borrower. In the second stage, we will use the LGD to find the suitable parametric model. In the third stage, we will use beta regression with different link functions for fitting the LGD data with covariates for two parameters for beta regression.

3.1 Loss-Given Defaults (LGD)

A variety of models in which LGD is subject to systematic risk can be found in the literature. Reference [28] proposed a model in which the LGD is normally distributed and influenced by the same systematic factor that drives the probability of default (PD). Reference [10] employs a lognormal distribution for the LGD. Other extensions include [29], choosing a probit transformation. References [30,31] employ a logit transformation. However, reference [6] used mortality approach to measure the percentage of bad and doubtful loan of corporate bonds that are recovered n months after the default date. The actuarial-based mortality approach is appropriate because the population sample is changing over time. The dataset of this study is obtained

from micro-data on defaulted bank loans of a private bank in Portugal, Banco Commercial Portugues (BCP). It consists of 10000 short-term loans granted to small and medium-sized companies from June 1995–December 2000 (66 months). They identified the LGD by the following: $LGD = \prod_{t=1}^{T} SPULB_t$, where $SPULB_t = 1 - SMRR_t$, and $SMRR_t = \frac{\sum_{i=1}^{m} Cashflow\ received_{it}}{\sum_{i=1}^{m} Loan\ outstanding_{it}}$ where $SPULB_t$ is a sample (weighted) percentage of unpaid loan balance at period t, $SMRR_t$ is a sample (weighted) marginal recovery rate at time t, i refers to each of the m loan balances outstanding in the sample, and t is the periods after default. They found that the cumulative average recovery is almost completed after 48 months. Moreover, the distribution of cumulative recovery rates is a bi-model distribution. Reference [32] used unsecured consumer loans or credit cards for one UK lender to compare linear regression and survival analysis models to predict LGD. The datasets were collected on 27000 personal loans from 1989–2004. There are two reasons to use survival analysis. Firstly, debts which are still being repaid cannot be included in the standard linear regression approach. Survival analysis models can treat such repayments as censored and easily include them in the model. Secondly, the recovery rate is not normally distributed and therefore modeling it using a linear regression violates the assumptions of linear regression models. The recovery rate is defined as: $RR = \frac{recovery\ amount}{default\ amount}$, where RR is a recovery rate and $LGD = 1 - RR$. The study compared linear regression with survival analysis models (proportional hazard models and accelerated failure time models for Weibull, log–logistic, gamma, and Cox model). The linear regression is better than survival models in single distribution models based on higher R-square, higher Spearman rank, and lower MSE. Reference [6] used LGD ratio as shown before to estimate the percentage of LGD after n-months of corporate bond default. However, reference [32] considered equation as shown before to estimate LGD ratio for personal loans. Our study constructs LGD ratio from previous equations to estimate LGD as shown in Eq. (1).

$$LGD = \frac{outstanding\ amount}{amount\ of\ borrowing} \tag{1}$$

The Basel II risk parameters are PD, LGD and exposure at default (EAD). The rate of expected credit loss (ECL) which is also known as the risk-weighted-asset of credit portfolio can be expressed as the product of PD and LGD. Therefore, LGD is one of the two determining factors of credit losses [33]. The ECL of credit portfolio can generally be represented as:

$$ECL = \sum_{i=1}^{m} PD_i * LGD_i * EAD_i \tag{2}$$

where PD_i is the probability of default of the ith borrower, $i = 1, 2, \ldots, m$, LGD_i is the loss given default, EAD_i is the exposure at default, and m is the total number of borrowers in the portfolio.

3.2 Parametric Distributions

Common parametric distributions for modeling the LGD data are considered. Tab. 1 provides the density function and survival function for the distributions considered in this study, which are Beta, Cauchy, Gamma, Gompertz, Logistic, Log-normal, Normal, and Weibull. These distributions are suitable for modeling skewed and heavy-tailed data, which are commonly displayed in LGD data. The empirical pdf for the LGD data in our study can be seen in Fig. 1 (pdf). The curve of empirical pdf indicates that the LGD data is skewed and heavy-tailed.

Table 1: Parametric models for LGD

| Distributions	Parameters	Probability density function $f(t)$	Survival function $S(t)$
Beta	Shape (α) Shape (β)	$\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ where $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$ $x \in [0,1]$	$I_x(\alpha,\beta) = \dfrac{B(x;\alpha,\beta)}{B(\alpha,\beta)}$ $B(x;\alpha,\beta)$ $\quad = \int_0^x t^{\alpha-1}(1-t)^{b-1}\,dt$
Cauchy	Mode (x_0) Scale (γ)	$\dfrac{1}{\pi\gamma\left[1+\left(\dfrac{x-x_0}{\gamma}\right)^2\right]}$	$\dfrac{1}{2} - \dfrac{1}{\pi}\arctan\left(\dfrac{x-x_0}{\gamma}\right)$
Gamma	Shape (α) Rate (β)	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{\frac{-t}{\beta}}$	$1 - \dfrac{\gamma(\alpha,\beta t)}{\Gamma(\alpha)}$ where $\gamma(\alpha,\beta t) = \int_0^{\beta t} z^{\alpha-1}e^{-z}.dz$
Gompertz	Shape (α) Rate(λ)	$\lambda e^{\alpha t}e^{\frac{-\lambda}{\alpha}\left(e^{\alpha t}-1\right)}$	$e^{\frac{-\lambda}{\alpha}\left(e^{\alpha t}-1\right)}$
Logistic	Mean (μ) Scale (s)	$\dfrac{e^{-\frac{x-\mu}{s}}}{s\left(1+e^{-\frac{x-\mu}{s}}\right)^2}$	$1 - \dfrac{1}{1+e^{-\frac{x-\mu}{s}}}$
Lognormal	Meanμ s.d.(σ)	$\dfrac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-(lnx-\mu)^2}{2\sigma^2}}$	$1 - \Phi\left(\dfrac{lnx-\mu}{\sigma}\right)$
Normal	Mean (μ) s.d.(σ)	$\dfrac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$1 - \Phi\left(\dfrac{x-\mu}{\sigma}\right)$
Weibull	Shape (α) Rate (β)	$\beta t^{\alpha-1}e^{-\beta t^\alpha}$	$e^{-\beta t^\alpha}$
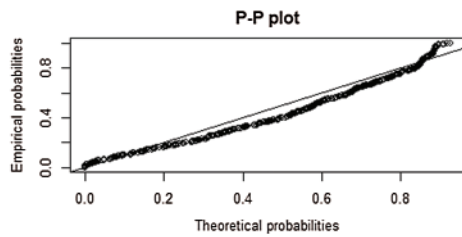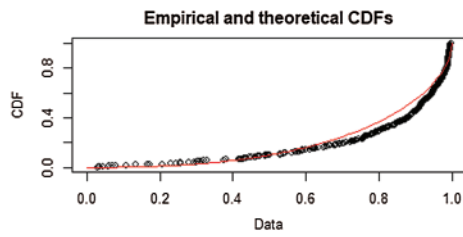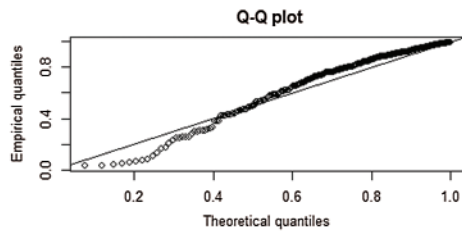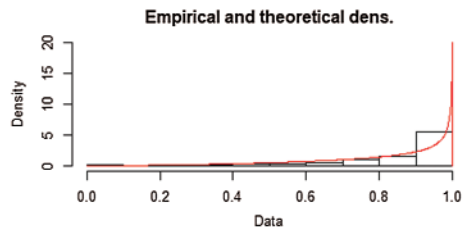
Beta distribution
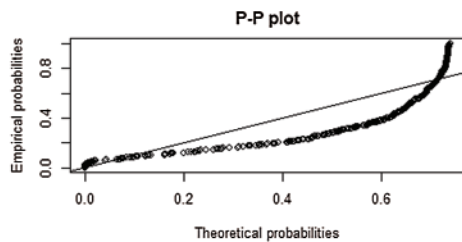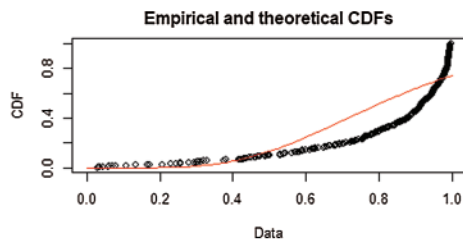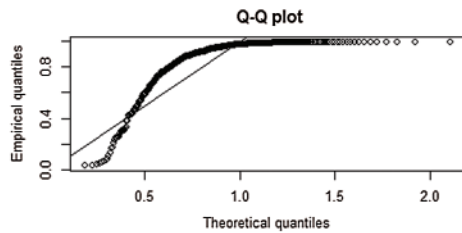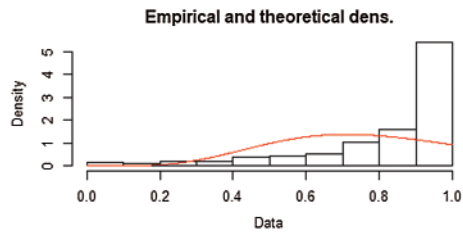
Gamma distribution
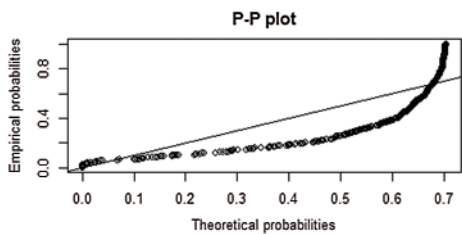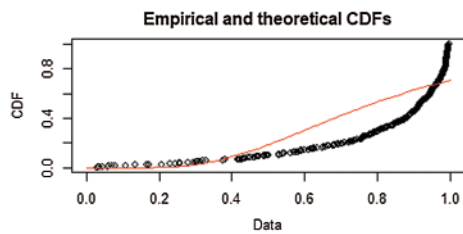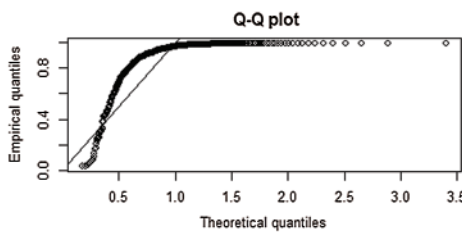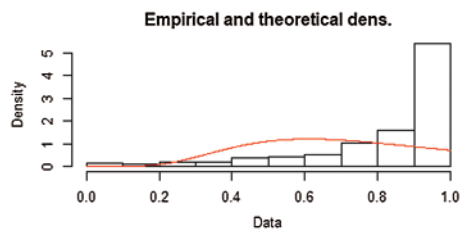
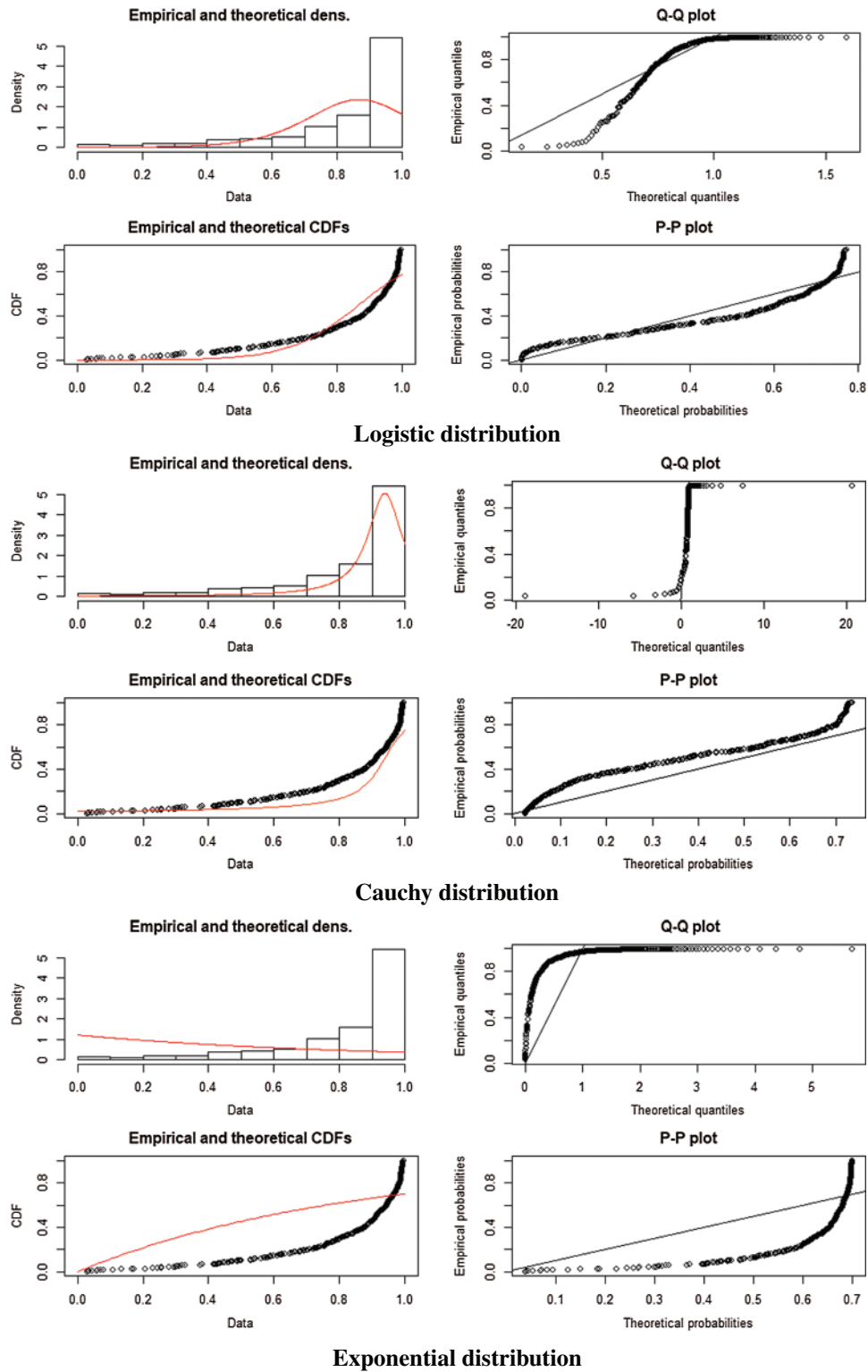**Normal distribution**

Figure 1: q–q plot, p–p plot, PDF and CDF. Beta distribution, gamma distribution, normal distribution, logistic distribution, Cauchy distribution, exponential distribution

Three types of accuracy criteria are used to choose the best model; Akaike information criterion (AIC), Bayesian information criterion (BIC), and Log-likelihood (LogL). BIC is depend on maximum likelihood estimates of the model parameters [34], which penalizes a sample data with larger size and number of parameters. The formula is defined as $BIC = -2\ell + kp$, where ℓ refers to the log likelihood of the estimated model, p refers to the number of parameters, and $k = \log n$. AIC is also depend on maximum likelihood estimates of model parameters [35], but penalizes a sample data with larger size. The formula is defined as $AIC = -2\ell + k * p$, where ℓ refers to the log likelihood of the estimated model, p refers to the number of parameters, and $k* = 2$.

3.3 Beta Regression

In our study we use beta regression for fitting the LGD data with covariates. The reason for fitting beta regression is that the distribution is well common to be adequate for modeling quantities bounded in the interval [0,1]. Based on the selection of parameters, the probability density function can be unimodal, J-shaped, U-shaped or uniform. It can be shown in the later section that beta distribution is the best model compared to other parametric distributions for fitting the sample data without covariates for our case study. Therefore, we consider several beta regressions, by using different link functions, for fitting the LGD data with covariates.

Let random variable Y follows to Beta distribution, $B(\alpha, \beta)$, where the parameters are $\alpha, \beta > 0$. The mean and variance of Y are defined as $E(Y) = \alpha/(\alpha + \beta)$ and $Var(Y) = \alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$ respectively. Reference [36] defined a regression structure of beta distribution as the followings. Let $\mu = \alpha/(\alpha + \beta)$ and $\phi = \alpha + \beta$, so that $\alpha = \mu\phi$ and $\beta = (1 - \mu)\phi$. The new parameterizations of Beta regression are $E(Y) = \mu$ and $Var(Y) = V(\mu)/(1 + \phi)$, where $V(\mu) = \mu(1 - \mu)$ for $Y \sim B(\mu, \phi)$ with $0 < \mu < 1$ and $\phi > 0$ since $\alpha, \beta > 0$. Parameter ϕ is known as precision parameter. Since a larger ϕ indicates as smaller variance for a fixed μ, $1/\phi$ can also be regarded as the dispersion parameter.

In our study, the precision parameter is modeled in a similar way as the mean parameter. Instead of having a fixed dispersion (fixed variance) we have a varying dispersion (varying variance). Therefore, a varying variance indicate a varying risks and it would be beneficial if significant risk factors (significant covariates) can be determined. It should be noted that variance is commonly used as one of the risk measures in finance area. The risk of loss in finance can be measured using confidence interval, for instance, the 95% confidence interval for loss can be measured as $\max(\mu \pm 2\sigma)$, where σ is the standard deviation.

Let $y = (y_1, \ldots, y_n)^T$ be a random samplewhere $y_i \sim B(\mu_i, \phi_i)$, $i = 1, \ldots, n$. The parameters, μ_i and ϕ_i, are assumed to satisfy the following functional relations:

$$g_1(\mu_i) = \eta_{1i} = f_1\left(x_i^T; \beta\right),$$

$$g_2(\phi_i) = \eta_{2i} = f_2\left(z_i^T; \theta\right), \tag{3}$$

where $\beta = (\beta_1, \ldots, \beta_k)^T$ and $\theta = (\theta_1, \ldots, \theta_h)^T$ are defined as vectors of unknown regression parameters that are assumed to be functionally independent, $\beta \in \mathbb{R}^k$ and $\theta \in \mathbb{R}^h$, $k + h < n$, η_{1i} and η_{2i} are predictors, and $x_{i1}, \ldots, x_{iq_1}, z_{i1}, \ldots, z_{iq_2}$ are observations on q_1 and q_2 known covariates which need not be exclusive.

A number of several link functions can be used for $g(.)$, such as logit specification which defined as $g(\mu) = \log(\mu/(1 - \mu))$, probit function $g(\mu) = \Phi^{-1}(\mu)$ where $\Phi(.)$ refer to the

standard normal distribution function, log function $g(\mu) = \log(\mu)$, complementary log–log function $g(\mu) = -\log(-\log(\mu))$, and Cauchy function $g(\mu) = \tan(\pi(\mu - 0.5))$. A rich discussion on the link functions can be explained in [37,38], or by referring to Chapter 7 in [39].

The log-likelihood function of Beta regression models defined as:

$$\ell(\beta, \theta) = \sum_{i=1}^{n} \ell_i(\mu_i, \phi_i) \tag{4}$$

where $\ell_i(\mu_i, \phi_i) = \log \Gamma(\phi_i) - \log \Gamma((1 - \mu_i)\phi_i) + (\mu_i \phi_i - 1) \log y_i + \{(1 - \mu_i)\phi_i - 1\} \log(1 - y_i)$, and $\mu_i = g_1^{-1}(\eta_{1i})$ and $\phi_i = g_2^{-1}(\eta_{2i})$ are functions of β and θ as defined in Eq. (3), respectively. The estimated parameters can be derived by maximizing the likelihood in (4). Further details on Beta regression can be found in [40].

In this study, we consider Beta regression with different link functions for fitting LGD data with covariates. We use R-squared to select the best regression model.

4 Empirical Results for LGD

A sample data based on the credit portfolio of a banking institution in Jordan is used in our study. The credit bank portfolio was obtained from January 2010–December 2014. The portfolio capacity is 4393, and the overall number of defaults during the 5-year period is 494. The sample size is the same as the number of default, which is 494, and the LGD data lies between interval [0,1]. For the case study, a borrower is declared as default if he is unable to pay cash installment in a period of 3 months.

The number of defaults *per annum* and the summary statistics for the LGD data are presented in Tab. 2. The maximum number of LGD is recorded in years 2010 and 2011. It can be observed that the highest mean of LGD is 97.7% with 0.025 standard deviation in 2010. The high frequency of LGD in 2010 and 2011 in as results of Jordanian economy had late response to financial crisis, which occurred in 2008. The Amman stock exchange (ASE) after 2009 starts decreasing steadily until 2012 [41,42]. Indicated that performance of Jordanian Banks sector had negative effect after global financial crisis such as, share prices decreasing and non-performing assets increasing. Therefore, for this reason we see the number of default increased as result of financial crisis that effect of ability of borrowers to pay their borrowing money. However, the lowest mean of LGD is 54.1% with 0.251 standard deviation in year 2014. Furthermore, the minimum LGD is 3.2% in year 2013 and the maximum LGD is 99.6% in year 2010. The R-package is used for fitting the sample data to the parametric models [43].

Tab. 3 provides the results of the fitted parametric distributions. Beta distribution is the preferable model because it has the highest LogL and the lowest AIC and BIC. Further comparison can be obtained from the q–q plot, p–p plot, empirical and theoretical PDF, and empirical and theoretical CDF shown in Fig. 1. It can be found that beta distribution shows a better fit compared to other parametric distributions. Therefore the mean of Beta distribution can be used as the estimate of LGD for the credit portfolio [44].

The results from Tab. 3 show that beta distribution is the best model for fitting the sample data. Therefore, we consider beta regression with different link functions in order to explore the internal and external financial variables which significantly affect the LGD.

The descriptive statistics of the explanatory variables are shown in Tab. 4.

Table 2: Summary statistics for LGD

Year	# of LGD	LGD			
		Mean	Std. Dev.	Min.	Max.
2010	137	0.9770	0.0248	0.8880	0.9963
2011	167	0.9164	0.0742	0.4516	0.9946
2012	67	0.7613	0.1647	0.1722	0.9833
2013	41	0.6274	0.2091	0.0324	0.8281
2014	82	0.5414	0.2507	0.2507	0.9472
Total	494				

Table 3: Parametric models

Family	Parameters	s.e.	LogL	AIC	BIC
Normal	Mean $= 0.826$ s.d $= 0.214$	0.0096 0.0068	60.30	-116.59	-108.19
Cauchy	Location $= 0.938$ Scale $= 0.063$	0.0045 0.0042	156.24	-308.47	-300.07
Logistic	Location $= 0.866$ Scale $= 0.106$	0.0081 0.0042	98.28	-192.56	-184.15
Beta	Shape1 $= 2.326$ Shape2 $= 0.559$	0.1641 0.0299	366.58	-729.15	-720.75
Exponential	Rate $= 1.211$	0.0545	-399.58	801.07	805.27
Gamma	Shape $= 7.184$ Rate $= 8.698$	0.4469 0.5605	-95.71	195.42	203.82
Weibull	Shape $= 5.110$ Scale $= 0.893$	0.2100 0.0080	64.68	-125.37	-116.96

Tab. 5 provides the parameter estimates and standard errors for Beta regressions, which are fitted using different link functions. All regression parameters are significant, at least at 0.10 level, except for beta regression with Cauchy link function where the estimate of LR is insignificant.

The estimates of GDP and Inf are highly significant for the mean (μ). GDP refer to a monetary measure of the market value of all final goods and services produced within the country's border in a specific period of time (typically 1 year). An increasing GDP means that the economy is expanding, and firms are producing and selling more products or services. When GDP declines, the economy is depicted as being in a downturn. During downturn, fewer goods and services are being sold, business profits turn down, unemployment rises and government tax collections fall.

An increasing inflation rate (Inf) indicates a sustained increment in the prices of goods and services over a specific period of time. Inflation referred to a reduction in the purchasing power

per unit of money. In other word, when inflation rate rises, each unit of currency buys fewer goods and services.

Table 4: Summary statistics for explanatory variables

Type of variables	Explanatory variables	n	mean	std. dev.
External financial variables (macroeconomic)	Gross domestic product (GDP)	494	0.0265	0.0018
	Inflation rate (Inf)	494	0.0343	0.0198
Internal financial variables (microeconomic)	Service pricing policy (SPP)	494	2.3083	1.7489
	Operation efficiency (OE)	494	0.9103	0.1806
	Liquidity ratio (LR)	494	0.1668	0.4652

Our results show that a decreasing GDP growth rate resulted in an increasing LGD, while an increasing inflation rate resulted in an increasing LGD. The results are expected in an adverse economic conditions, where GDP diminishes and inflation rate increases as the default frequencies increase and the asset prices decrease [44], and consequently, the recovery rates decrease. When an economic prosperity resumes, the situation reverses, indicating that the capital requirements under these conditions would swing wildly. In general, a lower LGD is favored for calculating the expected loss (EL) of a banking institution. The results show that lower LGD is obtained when the GDP is higher and the inflation rate is lower.

Service pricing policy (SPP) ratio is measured as operating expenses divided by total liability. This ratio measures the funding for operating expenses by the total liability. Higher SPP is resulted when the relative decrease in total liability is more than the decrease in operating expenses. For our case, higher SPP decreases the variance of LGD. It is implied here that larger decreases in total liability resulted in less variations of LGD among the default borrowers.

Operating efficiency ratio (OE) is measured as total operating expenses divided by total operating revenues. The increase in OE is caused by the larger decrease in total operating revenue relative to the decrease in total operating expenses. Our case shows that higher OE decreases the LGD variance. It can be indicated here that larger reduction in total operating revenue leads to less variations of LGD among the default borrowers.

Cash ratio (cash and cash equivalents divided by current liability) is generally a more conservative liquidity ratio measure of a company's ability to repay its short-term obligations, using only the most liquid of assets, such as cash on hand, cash equivalents (sometimes referred to as marketable securities) and demand deposits. This measure tells creditors the company's ability to pay all current liabilities immediately without having to sell or liquidate other assets. Higher LR is resulted when current liability has larger decrease than the decrease in cash and cash equivalents. Our results show that higher liquidity ratio (LR) indicates higher variance of LGD. It is implied here that a larger decrease in current liability resulted in more variations of LGD among the

Table 5: Beta regression with different link functions

Models	Parameters	Variables	Estimates	S.E	Z-value	
Beta regression (logit)	μ	(Intercept)	2.104	0.0439	47.90	****
		GDP*	−0.529	0.0324	−16.37	****
		Inf.*	0.879	0.0413	21.28	****
	$\varnothing$	(Intercept)	−0.596	0.2882	−2.07	**
		SPP	0.112	0.0341	3.28	***
		OE	2.850	0.3300	8.64	****
		LR	−0.188	0.1090	−1.72	*
Beta regression (loglog)	μ	(Intercept)	2.177	0.0416	52.34	****
		GDP*	−0.466	0.0276	−16.85	****
		Inf.*	0.801	0.0369	21.72	****
	$\varnothing$	(Intercept)	−0.597	0.2881	−2.07	**
		SPP	0.108	0.0341	3.18	***
		OE	2.850	0.3299	8.64	****
		LR	−0.185	0.1091	−1.69	*
Beta regression (probit)	μ	(Intercept)	1.222	0.0221	55.36	****
		GDP*	−0.306	0.0186	−16.51	****
		Inf.*	0.482	0.0229	21.02	****
	$\varnothing$	(Intercept)	−0.619	0.2883	−2.15	**
		SPP	0.120	0.0341	3.51	****
		OE	2.886	0.3302	8.74	****
		LR	−0.195	0.1086	−1.80	*
Beta regression (cloglog)	μ	(Intercept)	0.773	0.0169	45.73	****
		GDP*	−0.280	0.0172	−16.27	****
		Inf.*	0.404	0.0202	20.02	****
	$\varnothing$	(Intercept)	−0.632	0.2886	−2.19	**
		SPP	0.130	0.0342	3.81	****
		OE	2.917	0.3307	8.82	****
		LR	−0.205	0.1083	−1.90	*
Beta regression (Cauchy)	μ	(Intercept)	3.203	0.1604	19.96	****
		GDP*	−0.885	0.0613	−14.44	****
		Inf.*	1.747	0.1105	15.81	****
	$\varnothing$	(Intercept)	−0.531	0.2881	−1.84	*
		SPP	0.084	0.0340	2.47	**
		OE	2.755	0.3294	8.36	****
		LR	−0.163	0.1102	−1.48	
Beta regression (log)	μ	(Intercept)	−0.140	0.0048	−28.93	****
		GDP*	−0.116	0.0064	−18.27	****
		Inf.*	0.1403	0.0079	17.65	****
	$\varnothing$	(Intercept)	−0.791	0.2904	−2.72	***
		SPP	0.159	0.0344	4.64	****
		OE	3.222	0.3334	9.66	****
		LR	−0.240	0.1075	−2.23	**

Signif. codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1.

default cases. In general, lower variance of LGD is favored in terms of risk measures. The results show that lower variance is obtained when we have higher SPP, higher OE, and lower LR.

Table 6: Beta regression with R-squared, AIC and BIC

Beta regression	R-squared	Iteration		LogL	df	AIC	BIC
		(BFGS)	(Fisher scoring)				
Logit	0.50	24	2	560.7	7	−1107.4	−1078.0
Loglog	0.49	25	2	558.7	7	−1103.4	−1074.0
Probit	0.52	24	2	564.7	7	−1115.4	−1086.0
Cloglog	0.47	26	1	571.3	7	−1128.6	−1099.1
Cauchit	0.19	30	3	547.7	7	−1081.3	−1051.9
Log	0.27	30	1	600.3	7	−1186.6	−1157.2

Further results can be seen in Tab. 6, where the R-squared, log likelihood, AIC and BIC for each model are provided. Since beta regression with probit link function has the highest R-squared, while having accepted measurements for logL, AIC and BIC, this model is chosen as the best regression model for explaining the relationship between LGD and financial variables.

5 Conclusion

In the context of credit portfolio, LGD is the percentage of exposure that will be lost if a default occurs. Uncertainty with respect to the actual LGD is an important source of risks of credit portfolio, in addition to default risk. In this study, several parametric distributions were used to estimate LGD based on a sample of credit portfolio collected from a bank in Jordan from the period of 2010–2014. The results show that Beta distribution is the best parametric model for estimating LGD based on the following tests; logL, AIC and BIC. Several financial variables were then incorporated to the sample data to find the macro- and micro-economics determinants of LGD. The results show that Beta regression with probit link function has the highest R-squared with accepted measurements for logL, AIC and BIC. The results from beta regression models show that macroeconomic variables (GDP and Inflation rate) are significant for the mean parameter (μ), while microeconomic variables (SPP, OE and LR) are significant for the precision parameter ($\varnothing$). In particular, a decreasing GDP growth rate resulted in an increasing LGD, while an increasing inflation rate resulted in an increasing LGD. In terms of LGD risks, the variance (risks) of LGD are lower with lower SPP and lower OP, but higher with lower LR. We have proposed successfully the significant microeconomic variables which affect on precision parameter ($\varnothing$) and compared between six different link functions (Logit, loglog, probit, cloglog, cauchit, and log) by Beta regression.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   J. J. Jaber, N. Ismail and S. N. M. Ramli, "Credit risk assessment using survival analysis for progressive right-censored data: A case study in Jordan," *Journal of Internet Banking and Commerce*, vol. 22, no. 1, pp. 1–18, 2017.

[2]   B. Kollar and T. Kliestik, "Simulation approach in credit risk models," in *4th Int. Conf. on Applied Social Science, Information Engineering Research Institute, Advances in Education Research*, vol. 15, pp. 150–155, 2014.

[3]   E. I. Altman, "Measuring corporate bond mortality and performance," *Journal of Finance*, vol. 44, no. 4, pp. 909–922, 1989.

[4]   V. V. Acharya, S. T. Bharath and A. Srinivasan, *Understanding the Recovery Rates on Defaulted Securities*. UK: Centre for Economic Policy Research, 2003.

[5]   P. Nickell, W. Perraudin and S. Varotto, "Stability of rating transitions," *Journal of Banking & Finance*, vol. 24, no. 1–2, pp. 203–227, 2000.

[6]   J. Dermine and C. N. De Carvalho, "Bank loan losses-given-default: A case study," *Journal of Banking & Finance*, vol. 30, no. 4, pp. 1219–1243, 2006.

[7]   H. Z. Chen, "A new model for bank loan loss-given-default by leveraging time to recovery," *Journal of Credit Risk*, vol. 14, no. 3, pp. 1–29, 2018.

[8]   Y. Tanoue, A. Kawada and S. Yamashita, "Forecasting loss given default of bank loans with multi-stage model," *International Journal of Forecasting*, vol. 33, no. 2, pp. 513–522, 2017.

[9]   L. Thompson and T. Brandenburger, "Survival analysis methods to predict loss rates in credit card portfolios," in *SDSU Data Science Sym. Session*, South Dakota State University, USA, vol. 31, 2019.

[10] M. Pykhtin, "Recovery rates: Unexpected recovery risk," *Risk-London-Risk Magazine Limited*, vol. 16, no. 8, pp. 74–79, 2003.

[11] T. Bellotti and J. Crook, *Modelling and Estimating Loss Given Default for Credit Cards*. University of Edinburgh Business School, UK, CRC Working Paper, pp. 1–15, 2008.

[12] M. Bruche and C. González-Aguado, "Recovery rates, default probabilities, and the credit cycle," *Journal of Banking and Finance*, vol. 34, no. 4, pp. 754–764, 2010.

[13] X. Huang and C. W. Oosterlee, "Generalized beta regression models for random loss given default," *Journal of Credit Risk*, vol. 7, no. 4, pp. 45–70, 2011.

[14] T. Bellotti and J. Crook, "Loss given default models incorporating macroeconomic variables for credit cards," *International Journal of Forecasting*, vol. 28, no. 1, pp. 171–182, 2012.

[15] R. Calabrese, "Estimating bank loans loss given default by generalized additive models," *Technical Report University College Dublin Geary Institute, Working Paper*, vol. 201224, pp. 1–17, 2012.

[16] G. Loterman, I. Brown, D. Martens, C. Mues and B. Baesens, "Benchmarking regression algorithms for loss given default modeling," *International Journal of Forecasting*, vol. 28, no. 1, pp. 161–170, 2012.

[17] P. Li, M. Qi, X. Zhang and X. Zhao, "Further investigation of parametric loss given default modeling," *Journal of Credit Risk, Forthcoming*, vol. 12, no. 4, pp. 17–47, 2016.

[18] A. Hamerle, M. Knapp and N. Wildenauer, *Modelling Loss Given Default: A "Point in Time" Approach*, B. Engelmann and R. Rauhmeier, New York: Springer, pp.127–142, 2006.

[19] S. Krüger and D. Rösch, "Downturn LGD modeling using quantile regression," *Journal of Banking & Finance*, vol. 79, pp. 42–56, 2017.

[20] BCBS, "Base II: International convergence of capital measurement and capital standards: A revised framework," in *Basel Committee on Banking Supervision*. Basel: Bank for International Settlements, 2004.

[21] BCBS, "Basel II: Background note on LGD quantification," in *Basel Committee on Banking Supervision*. Basel: Bank for International Settlements, 2004.

[22] BCBS, Basel II: Guidance on paragraph 468 of the framework document. Basel Committee on Banking Supervision. Basel: Bank for International Settlements, 2005.

[23] E. I. Altman, B. Brady, A. Resti and A. Sironi, "The link between default and recovery rates: Theory, empirical evidence, and implications," *Journal of Business*, vol. 78, no. 6, pp. 2203–2228, 2005.

[24] S. Figlewski, H. Frydman and W. Liang, "Modeling the effect of macroeconomic factors on corporate default and credit rating transitions," *International Review of Economics and Finance*, vol. 21, no. 1, pp. 87–105, 2012.

[25] V. V. Acharya, S. T. Bharath and A. Srinivasan, "Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries," *Journal of Financial Economics*, vol. 85, no. 3, pp. 787–821, 2007.

[26] S. Caselli, S. Gatti and F. Querci, "The sensitivity of the loss given default rate to systematic risk: New empirical evidence on bank loans," *Journal of Financial Services Research*, vol. 34, no. 1, pp. 1–34, 2008.

[27] P. Grippa, S. Iannotti and F. Leandri, "Recovery rates in the banking industry: Stylised facts emerging from Italian experience," in *Recovery Risk: The Next Challenge in Credit Risk Management*. London: Riskbooks, pp. 121–141, 2005.

[28] J. Frye, "Depressing recoveries," *Risk-London-Risk Magazine Limited*, vol. 13, no. 11, pp. 108–111, 2000.

[29] L. Andersen and J. Sidenius, "Extensions to the Gaussian copula: Random recovery and random factor loadings," *Journal of Credit Risk*, vol. 1, no. 1, pp. 5, 2004.

[30] K. Düllmann and M. Gehde-Trapp, "Systematic risk in recovery rates: An empirical analysis of US corporate credit exposures," in *Discussion Paper, Banking and Financial Supervision*, Deutsche Bundesbank Frankfurt, Germany, pp. 1–44, 2004.

[31] D. Rösch and H. Scheule, *A Multi-Factor Approach for Systematic Default and Recovery Risk, in the Basel II Risk Parameters*. Berlin: Springer, 105–125, 2006.

[32] J. Zhang and L. C. Thomas, "Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD," *International Journal of Forecasting*, vol. 28, no. 1, pp. 204–215, 2012.

[33] M. Qi and X. Zhao, "Comparison of modeling methods for loss given default," *Journal of Banking & Finance*, vol. 35, no. 11, pp. 2842–2855, 2011.

[34] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[35] H. Akaike, "Fitting autoregressive models for prediction," *Annals of the Institute of Statistical Mathematics*, vol. 21, no. 1, pp. 243–247, 1969.

[36] S. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions," *Journal of Applied Statistics*, vol. 31, no. 7, pp. 799–815, 2004.

[37] P. McCullagh and J. Nelder, *Generalized Linear Models*. London: Chapman and Hall, 1989.

[38] A. V. Rocha and A. B. Simas, "Influence diagnostics in a general class of beta regression models," *Test*, vol. 20, no. 1, pp. 95–119, 2011.

[39] A. C. Atkinson, *Plots, Transformation and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford: Clarendon Press, 1985.

[40] A. B. Simas, W. Barreto-Souza and A. V. Rocha, "Improved estimators for a general class of beta regression models," *Computational Statistics and Data Analysis*, vol. 54, no. 2, pp. 348–366, 2010.

[41] A. A. A. Sharabati and A. N. I. Nour, "Impact of global financial crisis on Amman Stock Exchange (ASE) market," *Journal of Baghdad College of Economic sciences University*, vol. 2014, no. 4, pp. 540–558, 2014.

[42] B. A. Dalaien, "Impact of global financial crisis on banking sector of India and Jordan," *Academic Journal of Economic Studies*, vol. 2, no. 1, pp. 79–95, 2016.

[43] N. Belgorodski, M. Greiner, K. Tolksdorf, K. Schueller, M. Flor *et al.,* "Rriskdistributions: fitting distributions to given data or known quantiles," *The Comprehensive R Archive Network*, pp. 1–68, 2017.

[44] G. Van Vuuren, R. De Jongh and T. Verster, "The impact of PD-LGD correlation on expected loss and economic capital," *International Business and Economics Research Journal*, vol. 16, no. 3, pp. 157–170, 2017.