Effect of Firm Age in Expected Loss Estimation

for Small Sized Firms

Kenzo Ogi†
Risk Management Department
Japan Finance Corporation, Tokyo, Japan
Tel: (+81) 3-3270-1408, Email: ogi-k@jfc.go.jp

Masahiro Toshiro
Risk Management Department
Japan Finance Corporation, Tokyo, Japan
Tel: (+81) 3-3270-1408, Email: toshiro-m@jfc.go.jp

Norio Hibiki
Faculty of Science and Technology
Keio University, Yokohama, Japan
Tel: (+81) 45-566-1635, Email: hibiki@ae.keio.ac.jp

Abstract. In the banking industry, Expected Loss (EL) is one of the most important indicator for calculating credit cost, allowance for doubtful accounts, loan interest rates, and so on. EL is calculated by multiplying Probability of Default (PD) and Loss Given Default (LGD). Hence, there are a number of studies concerning prediction models of PD or LGD. Most of Japanese banks have utilized prediction models of PD, whereas few Japanese banks have employed prediction models of LGD. This is because of the fact that the collateral coverage in lending for small sized firms is almost 100%. Consequently, most Japanese banks do not consider the correlations between PD and LGD of an individual borrower. If such a positive correlation exists, the EL may be seriously underestimated when loan portfolio structures change and PD increases. According to some previous studies on credit risk for small sized firms, it is very likely that the correlations in small sized firms are positive. However, to the best of our knowledge, there are no studies which examine these positive correlations of small sized firms. In this paper, we analyze such correlations by using a data set of approximate 630,000 Japanese small sized firms for the period from 2004 to 2012, owned by Japan Finance Corporation. The results of our analyses show as follows; i) the correlations of collateralized loans are low, whereas those of uncollateralized loans are rather high and positive, ii) the observed correlation is spurious in fact, and the common factor is firm age, and iii) EL is able to be approximately estimated by utilizing the single factor model that uses the firm age as the variable instead of PD and LGD.

Keywords: Expected Loss, Probability of Default, Loss Given Default, credit risk, firm age

1. INTRODUCTION

Banks need to estimate Expected Loss (EL) for calculating credit cost, allowance for doubtful accounts, loan interest rates, and so on. It is one of the most important indicator for not only banks but also Micro Business and Individual Unit of Japan Finance Corporation (JFC-Micro), a policy-based financial institution that aims to contribute to the promotion of small sized firms and solo proprietors.

EL is calculated by multiplying Probability of Default (PD) and Loss Given Default (LGD), and there are many studies concerning prediction models of PD or LGD. The prediction of PD for potential and existing borrowers has been the important subject during the past few decades, whereas the prediction of LGD has become popular just recently. The impact of LGD has been studied less than PD. Furthermore, few Japanese banks need to employ prediction models of LGD because of the fact that collateral coverage in lending for small sized firms is almost 100%. Therefore, most Japanese banks do not

Table 1 : Data Set

Implementation of loan (FY)	Borrower's default year (FY)	Cumulative LGD of 1 year after the default (FY)	Cumulative LGD of 2 years after the default (FY)	Cumulative LGD of 3 years after the default (FY)	Collateralized loan (the number of firms)	Uncollateralized loan (the number of firms)
2012	2013				18,464	58,503
2011	2012	2013			9,429	57,402
2010	2011	2012	2013		12,603	77,700
2009	2010	2011	2012	2013	13,787	86,221
2008	2009	2010	2011	2012	11,421	75,474
2007	2008	2009	2010	2011	10,176	63,670
2006	2007	2008	2009	2010	10,638	45,608
2005	2006	2007	2008	2009	13,450	30,784
2004	2005	2006	2007	2008	14,680	19,108
Total					114,648	514,470

consider the correlation between PD and LGD of an individual borrower. If such a positive correlation exists, the EL may be seriously underestimated when loan portfolio structures change and PD increases.

There are various studies concerning the correlation between PD and LGD of loan portfolio. Altman *et al*. (2002) and Moody's (2011) show that there is a positive correlation between a default rate (DR) and LGD, where the DR is observed value of PD. Also, they analyze dependence of DR and LGD on common macroeconomic factors such as GDP growth, unemployment rate, consumer spending, the default rate itself, and so on. On the other hand, Hart and Felsovalyi(1998) and Witzany(2011) report that the correlation is not significant at 5% level. However, above studies do not focus on the correlation for an individual borrower.

Grunert and Weber (2009) present a negative relationship between PD and LGD of an individual borrower by analyzing 120 companies that borrowed from one large German bank, and points out that the correlations may lead to underestimation of the credit risk calculated by the credit risk model. However, the number of data is only 120 companies that include large companies, and they haven't analyzed EL behavior in case that loan portfolio structures change and PD increases. Kawada and Yamashita (2012) analyze about 90,000 small and medium sized firms that borrowed from three Japanese banks, and report that the correlation between credit score of borrower and LGD is negative contrary to their expectation.

As mentioned above, few studies have focused on the correlation between PD and LGD of an individual borrower. Furthermore, to the best of our knowledge, whether the correlation in small sized firms is positive or negative has not been clarified.

According to some previous studies on credit risk of small sized firms, it is very likely that the correlations of small sized firms are positive. Ogi, Toshiro and Hibiki (2014) note that the firm age is a proxy variable of the private asset of the business owner, and DR tends to decrease as a firm becomes older. In addition, Ogi *et al*. (2015) report that LGD tends to decrease as a firm becomes older, and the firm age is one of the important factors of prediction model of LGD, as well as PD. Previous studies by our group have shown that the firm age as a proxy variable of private asset of the business owner impacts on both PD and LGD. Thus, we expect that the correlations between PD and LGD of small sized firms must be positive.

In this paper, we focus the analysis on the correlation between PD and LGD of an individual borrower calculated by using a data set of approximate 630,000 Japanese small sized firms with 20 or less employees for the period from 2004 to 2012, owned by Japan Finance Corporation. Then, we find that the positive correlation exists between PD and LGD, and the firm age is the common factor. Our main findings are as follows.
i) The correlations of collateralized loans are low, whereas those of uncollateralized loans are rather high and positive.
ii) The firm age is the common factor of describing PD and LGD, and therefore the observed correlation is spurious.
iii) EL can be approximately estimated by utilizing the single factor model that includes the firm age as one factor instead of PD and LGD.

The paper is organized as follows. Section 2 shows empirical results concerning the correlations between DR and LGD. In Section 3, we test a hypothesis that the firm age is a common factor. Section 4 empirically examines the accuracy of the single factor model. Section 5 provides our conclusion.

2. EMPIRICAL RESULTS CONCERNING THE CORRELATIONS BETWEEN DR AND LGD

We examine the positive correlation between PD and LGD of an individual borrower in small sized firms. In this paper, we use a data set of approximate 630,000 small sized firms to which JFC provided loans from 2004 to 2012.

Table 2 : Analytical procedure

	Firm (No.)	0001	0002	0003	0004	· · ·	999	1000
1)	Credit score	56.4	82.5	45.8	32.6	· · ·	76.1	63.2
	Status	Default	Non-Default	Non-Default	Default	· · ·	Non-Default	Non-Default

↓

	Group	1		2		· · ·	16	
2)	The number of Default firms	525		525		· · ·	525	

↓

	Group	1		2		· · ·	16	
	Threshold of credit score	Score > X_1		$X_1 \geqq$ Score > X_2		· · ·	Score > X_{16}	
3)	The number of Default firms	525		525		· · ·	525	
	The number of Non-Default firms	73,750		49,925		· · ·	6,079	

↓

	Group	1		2		· · ·	16	
4)	DR	••		••		· · ·	••	
	LGD	••		••		· · ·	••	

↓

5)   LGD=α+β（DR）

Also, since Ogi *et. al*(2015) show that factors influencing the LGDs vary according to collateral coverage, we divide the sample into three parts by collateral coverage level : (i) 0%(No)coverage, (ii) 100%(full)coverage, (iii) other(more than 0% and less than 100%). We analyze the cases of (i) and (ii) in this paper. Moreover, we test cumulative LGD for one to three years elapsed from the borrower's default, because the work-out recovery processes often take few years. As a result of analyses, we confirm that the correlations of collateralized loans are low, whereas the correlations of uncollateralized loans are high.

2.1 Data Set and Definition of DR and LGD

Table 1 shows the data set. We use information of 629,118 firms that borrowed from JFC-Micro from 2004 to 2012. The number of collateralized loans is 114,648, and that of uncollateralized loans is 514,470.

JFC-Micro provides business loans for micro/small sized firms. Approximately 90% of borrowers run the businesses with nine or less employees. About a quarter of all small sized firms in Japan get loan from JFC-Micro, but we need to pay attention that the sample consists of only firms for which JFC-Micro made loans.

The DR at time t is calculated as follows:

$$DR(t) = \frac{D(t)}{ND(t) + D(t)} , \qquad (1)$$

where *ND(t)* is the number of the non-default firms at time *t*, and *D(t)* is the number of the default firms at time *t*.

LGD is expressed by one minus Recovery Rate (RR) which is defined as the fraction of getting paid back. The LGD at time *t* is calculated as follows:

$$LGD_{i,\tau}(t) = 1 - RR_{i,\tau}(t) = 1 - \frac{CF_{i,\tau}(t)}{EAD_i(t)} \qquad (2)$$

where i is each firm, $LGD_{i,\tau}(t)$ is cumulative LGD of τ years after default, $CF_{i,\tau}(t)$ is the sum of the payback for τ years, and $EAD_i(t)$ is exposure at default at time t.

2.2 Analytical procedure

In this chapter, we use the data set of the period from 2007 to 2011 because samples of uncollateralized loans from 2004 to 2006 are insufficient to analyze. The procedure of the analysis is as follows:
1) We calculate credit score of each firm by using credit scoring model made by JFC-Micro.
2) After sorting default firms in descending order of their credit score, we divide them into 16 groups of the equal size.
3) We calculate the threshold based on the minimum and maximum scores of each group, and allocate non-default firms to 16 groups in accordance with the threshold scores.
4) We calculate DR and LGD of each category.
5) We examine the correlations between DR and LGD by using ordinary least squares (OLS) analysis.

2.3 Results

Table 3 shows the adjusted coefficient of determination through regression analysis for the sample of the collateralized loans from 2007 to 2011, and subsamples of each year. The correlations of most groups are low because the adjusted R-squared is equal to or less than 0.1. Figure 1 shows the scatter plot to examine the correlations between PD and LGD of the collateralized loans. We use

Table 3 : Adjusted coefficient of determination of the collateralized loans in each year

Implementation of loan (FY)	Borrower's default year (FY)	Cumulative LGD of 1 year after the default (FY)	Cumulative LGD of 2 years after the default (FY)	Cumulative LGD of 3 years after the default (FY)
2007-2011	2008-2012	0.02	0.07	0.02
2011	2012	-0.08		
2010	2011	0.00	0.00	
2009	2010	0.00	0.01	0.00
2008	2009	0.01	0.00	0.02
2007	2008	0.01	0.04	0.03

Table 4 : Adjusted coefficient of determination of the uncollateralized loans in each year

Implementation of loan (FY)	Borrower's default year (FY)	Cumulative LGD of 1 year after the default (FY)	Cumulative LGD of 2 years after the default (FY)	Cumulative LGD of 3 years after the default (FY)
2007-2011	2008-2012	0.83	0.82	0.82
2011	2012	0.33		
2010	2011	0.27	0.41	
2009	2010	0.48	0.42	0.44
2008	2009	0.17	0.43	0.60
2007	2008	0.65	0.76	0.76

the cumulative LGD of 3 years elapsed from the borrower's default. We cannot find the positive correlation between DR and LGD of the collateralized loans in the plot.
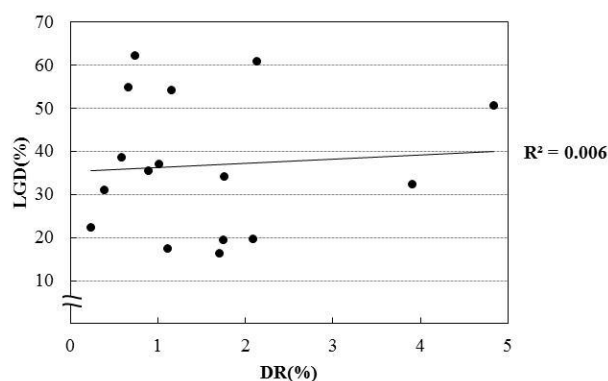

Figure 1: The correlations between PDs and LGDs of the collateralized loans

Table 4 shows the adjusted coefficients of determination for the uncollateralized loans of each year. The adjusted R-squared values of 2007-2011 samples are around 0.82. Those of each year are between 0.41 and 0.76 for the cumulative LGD of 2 -3 years.

In addition, Figure 2 shows the scatter plot between DR and LGD of the uncollateralized loans. The result shows that the correlations are high because the adjusted R-squared is about 0.82. The positive correlations can be found for the uncollateralized loans. The RR of uncollateralized loans is influenced by the size of assets at default. As the PD of the firm is lower, the remaining asset

at default becomes larger. This finding is consistent with the practical perspective.

$y = 0.9339x + 87.877$
$R^2 = 0.8171$
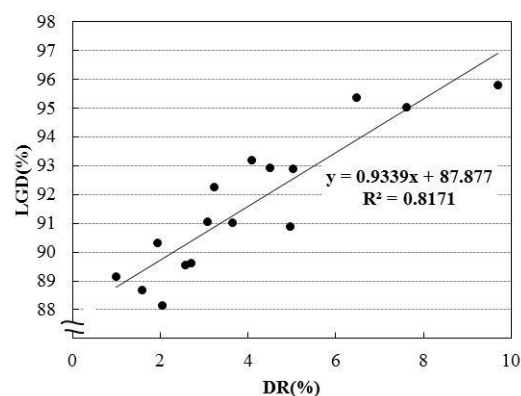
Figure 2: The correlations between PDs and LGDs of the uncollateralized loans

3. TEST HYPOTHESIS THAT THE COMMON FACTOR IS FIRM AGE

We find out the positive correlations between PD and LGD of an individual borrower of uncollateralized loans. Recently, banks have increased uncollateralized loans, and loan portfolio structures change. The EL may be seriously underestimated when PD increase, and therefore Japanese banks should consider such correlations to estimate the appropriate EL. It is effective to incorporate the common factor into both prediction models of PD and LGD in order to consider the correlations.

The common factor is a variable that has the high correlation with both PD and LGD. It is presumed that there are various variable candidates such as capital-to-

Figure 4: Relationship between LGD and firm age each number of years elapsed from the borrower's default
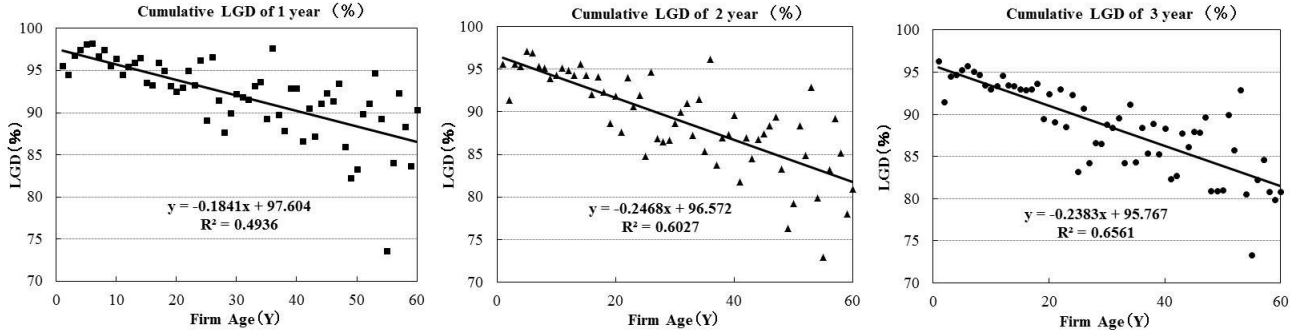
asset ratio and net assets amount. Ogi *et al*. (2014, 2015) show that the firm age is a proxy variable of owner's private assets, and that is one of the important factors of prediction model of PD and LGD. In small sized firms, since owner's private assets have a strong impact on running their business, we expect that the common factor is the firm age.

In Section 3, we test the hypothesis that the common factor is the firm age. First, we formulate the DR and LGD measured every firm age bracket by using the *n*-th order polynomial function. Next, we attempt to examine the correlation between residuals of PD and LGD measured every firm age bracket. This residual is the difference between an observed value and its prediction calculated by these formulas. If the correlations between residuals are low, this hypothesis is supported.

3.1 Formulation of the DR measured every firm age bracket

Ogi *et al*. (2014) reveal that the DR measured every firm age bracket is expressed by the cubic function, and they examine the robustness concerning some firm characteristics such as industry, firm size, and rating classification, but not collateral coverages.   In Section 3.1, we attempt to formulate the DR measured every firm age bracket in uncollateralized loans by using the *n*-th order polynomial function. Using OLS method, the function is set as follows:

$$p_g = a_0 + \sum_{k=1}^{n} \beta_k g^k \quad (g = 1, \cdots, G), \tag{3}$$

where g is the firm age, G is the maximum firm age, p_g is DR in g years, β_k is the OLS coefficient of the k-th power of the firm age, and a_0 is the constant term.

Table 5 reports adjusted R-squared for different-order polynomial functions. Adjusted R-squared does not increase even when the polynomial order is higher than three. As shown in Figure 3, we find that DR measured every firm age bracket of uncollateralized loans is able to

be expressed by the cubic function as well. These findings are consistent with our previous study.

Table 5: Adjusted R-squared for different-order polynomial functions in the DR measured every the firm age bracket

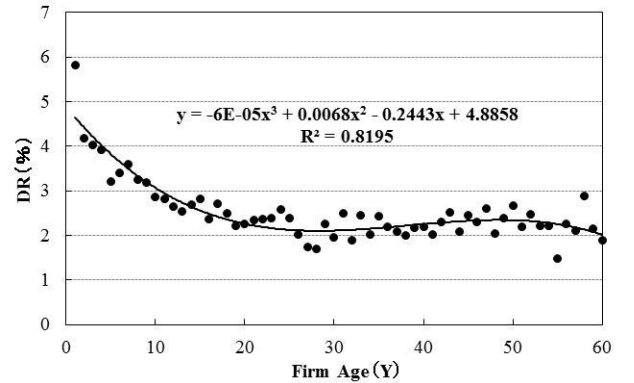| order(n)	n=1	n=2	n=3	n=4	n=5	n=6
Adjusted R^2	0.396	0.696	0.820	0.829	0.842	0.859

Figure 3: Relationship between DR and firm age of uncollateralized loans

3.2 Formulation of the LGD measured every firm age bracket

Ogi *et al*. (2015) show that the firm age is effective as a variable of prediction model of LGD. However, they did not formulate the LGD concerning firm age. In this section, we attempt to formulate it on uncollateralized loans utilizing as the same method as Section 3.1.

We need to define the number of years from default in order to calculate the LGD measured every firm age bracket. Due to the fact that bank loans are usually non-tradable, debt-collecting methods are mainly work-out recovery process that often takes a few years. For this

reason, we test cumulative LGD for one to three years elapsed from the borrower's default by using the data set from 2007 to 2009. The result is shown in Figure 4. The adjusted R-squared value of cumulative LGD of three years is the highest. Based on this result, we carry out our analysis by using the data set of cumulative LGD of three years elapsed from the borrower's default.

Table 6 reports adjusted R-squared values for different-order polynomial functions. The result shows the first-order is enough to express the LGD because the adjusted R-squared value does not increase for the higher order. We find that LGD measured every firm age bracket of uncollateralized loans is expressed by the linear function.

Table 6: The adjusted R-squared for different-order polynomial functions in LGD

order(n)	n=1	n=2	n=3	n=4	n=5	n=6
Adjusted R^2	0.656	0.656	0.657	0.680	0.680	0.684

3.3 Result of analysis

The DR and LGD can be formulated as functions of firm age by formulas (4) and (5), respectively. If the correlation between $\varepsilon_{DR,x}$ and $\varepsilon_{LGD,x}$ is low, it is more likely that the firm age is the common factor.

$$y_{DR,x} = -6 \times 10^{-5}x^3 + 0.0068x^2 - 0.2443x + 4.8858 + \varepsilon_{DR,x} \tag{4}$$

$$y_{LGD,x} = -0.2383x + 95.767 + \varepsilon_{LGD,x} \tag{5}$$
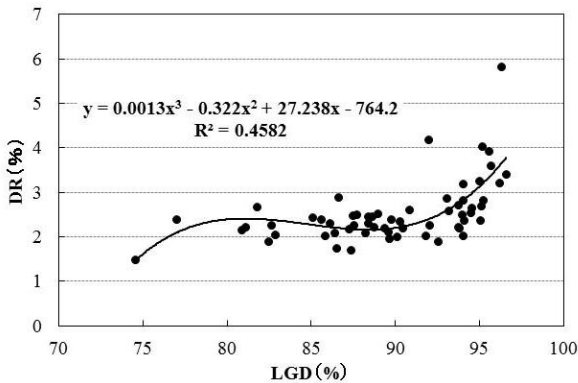
Figure 5: The relationship between DR and LGD measured every firm age bracket

First, Figure 5 shows the relationship between DR and LGD measured every firm age bracket. These relations can be represented by the cubic function, and its adjusted R-squared is 0.4582. This seems to be sufficient level.

Next, Figure 6 shows the relationship between residuals. We find that the correlation is low. We confirm that the observed correlation between PD and LGD is spurious and the common factor is the firm age.
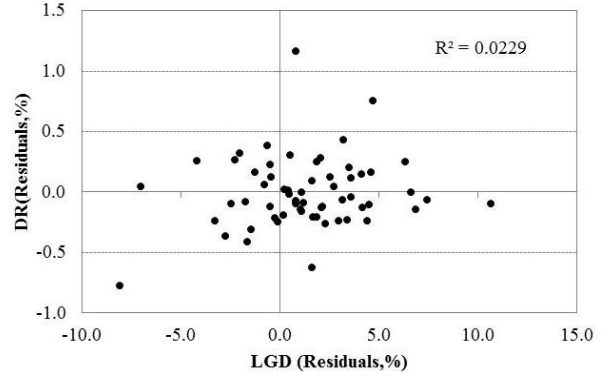
Figure 6: Relationship between residuals of DR and LGD

4. SINGLE FACTOR MODEL OF CALCULATING EXPECTED LOSS

The firm age is the common factor between PD and LGD of an individual borrower. This fact gives us the possibility to approximately estimate the EL by using firm age instead of PD and LGD. We make the single factor model to estimate EL, and attempt to validate the robustness of the model.

As a result of the analysis, when the average observed rate of loss of 20,000 firms is 2.2%, the average absolute difference between observed rate and its prediction is about 0.116%, and the standard deviation is about 0.090%. These results support our hypothesis that it is likely to approximately estimate the EL by using only firm age from a practical perspective.

4.1 Formulation of EL as a function of firm age

The EL is shown by formula (6), and therefore we attempt to express the EL by a quartic function because the DR is expressed by a cubic function (7) and the LGD is expressed by a linear function (8) as shown in the previous section.

$$EL = PD \times LGD \tag{6}$$

$$PD(3) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_1 x^3 = \sum_{i=0}^{3} \alpha_i x^i \tag{7}$$

$$LGD(1) = \beta_0 + \beta_1 x = \sum_{j=0}^{1} \beta_j x^j \tag{8}$$

Table 7 : Examples of sample selected in random order

No	Firm Age	Loan volume	Loan's balance after 1 year	EAD	Recovery amount
1	26	11,000	8,800		
2	11	15,000	12,000	12,000	1,000
3	35	5,000	4,000		
4	21	8,000	6,400		
5	18	35,000	28,000		
6	6	2,000	1,800	1,800	0
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
100	51	25,000	15,000		
Total		1,000,000	800,000	120,000	12,000

Note: This table shows the example of the 100-firm sample.

Necessarily, the EL is expressed by the quartic function (9), where x is the firm age, i or j is n-th power of the firm age.

$$\text{EL}(4) = \sum_{i=0}^{3} \sum_{j=0}^{1} \alpha_i x^i \, \beta_j x^j = \sum_{k=0}^{4} \eta_k \, x^k,$$

$$\text{where,} \quad \eta_k = \sum_{t=\max(k-1,0)}^{\min(k,3)} \alpha_t \, \beta_{k-t} \qquad (9)$$

Empirical result is shown in Table 8. As we expected, the adjusted R-squared value of the quartic function is the highest, or 0.846.

Table 8: Adjusted R^2 for different-order polynomial functions in EL

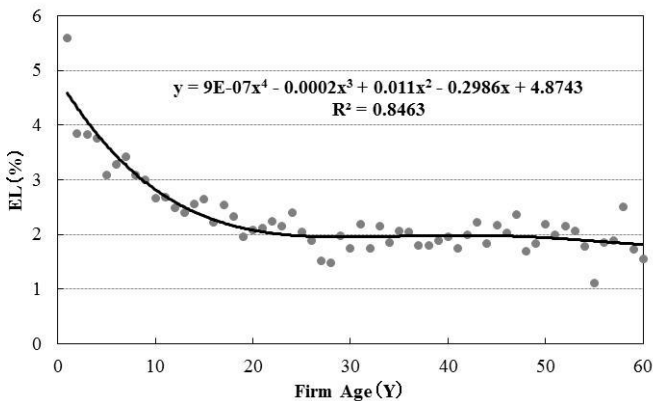| order(n)	n=1	n=2	n=3	n=4
Adjusted R^2	0.482	0.737	0.840	0.846

Figure 7: Relationship between EL and firm age

4.2 Evaluation of EL estimation model

As shown above, it is clear that we can approximately estimate the EL by using the model which single factor is firm age. We verify practical effectiveness of this model. First, we select samples in random order from the data set from 2004 to 2009 as shown in Table 7, and prepare the samples which size is 100, 300, 500, 1000, 3000, 5000, 10000 and 20000 firms. We generate 100 sets of samples. Sampling method is a single random sampling of SAS/STAT ® . Finally, we calculate the real loss by plugging observed values such as EAD, recovery amount and loan's balance after 1 year into formula (10).

$$\text{Real loss} = (\text{EAD} - \text{recovery amount}) \, / $$
$$\text{loan's balance after 1 year} \qquad (10)$$

Second, we calculate the EL by using the single factor model shown in Figure 7, and calculate the average/standard deviation of absolute difference between the real loss and the predicted loss by the model. The results are shown in Figure 8.

As the sample size is larger, the average absolute difference decreases. The average absolute difference of the 500 firm samples is about 0.64%. This is higher than our expectation because the average real loss is approximately 2.2%.

However, in case of the 20,000 firm samples, the average absolute difference is about 0.116%, and the standard deviation is about 0.09%. This result gives us the possibility to approximately estimate the EL by using only firm age from a practical perspective.
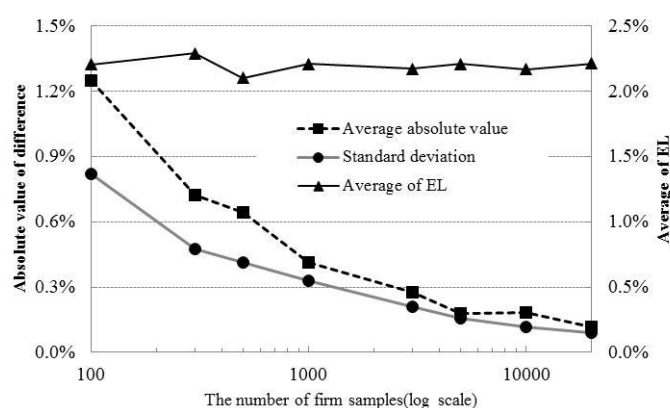
Figure8: Average/standard deviation of absolute difference between EL and actual value of loss

5. Conclusions

For most Japanese banks, it is not necessary to consider LGD from a practical perspective, because collateral coverage in lending for small sized firms is almost 100%. Therefore, they do not consider the correlations between PD and LGD of an individual borrower. Also, there are few studies concerning the correlation, compared with the prediction models of PD or LGD. However, as Japanese banks recently have increased uncollateralized loans, they have become taking more interest in the correlation year by year. Thus, we analyzed the correlation between PD and LGD for an individual borrower by using a data set of approximate 630,000 Japanese small sized firms with 20 or less employees, which is owned by the Japan Finance Corporation.

The first key finding of this paper is that the correlations between PD and LGD of collateralized loans are low, whereas those of uncollateralized loans are rather high and positive. The second key finding is that the observed correlation is spurious and the common factor is the firm age. The third key finding is that EL is approximately estimated utilizing the model which single factor is the firm age instead of PD and LGD.

According to above key findings, we can provide suggestive evidence for the following points of banking operation in practice.
1) Banks that do not consider the correlations between PD and LGD are likely to underestimate EL.
2) Considering the correlation, it is effective to incorporate the firm age factor into both prediction models of PD and LGD.
3) It is difficult for small sized financial institutions or business companies to estimate the EL because their data

is insufficient to estimate appropriately. Hence, the fact that EL can be estimated approximately by utilizing the single factor model gives them an alternative on calculating their EL.

We expect that our study enhances the potential for a practical use on sound banking. However, we have two aspects which we pay attention to in our results. First, it is possible that these findings are weakly biased, because the data set consists of only the firms for which JFC financed. In order to improve the estimation accuracy, we might need to update the model using another more refined data set. Second, the time-series analysis cannot be conducted in this paper because of the short span of the data. These are our future research.

REFERENCES

Altman E.I., A. Resti and A. Sironi．(2002) The link between default and recovery rates: effects on the procyclicality of regulatory capital ratios, Bank for International Settlement, *BIS Working Papers* No 113.

Grunert,J., Weber,M.(2009) Recovery rates of commercial lending : empirical evidence for German companies, *Journal of Banking and Finance* 33, 505-513.

Hurt, L. and A. Felsovalyi(1998) Measuring loss on Latin A merican defaulted bank loans, A 27-year study of 27 countries, *Journal of Lending and Credit Risk Management*, 81(2), 41-46.

Kawada,A. and Yamashita,S.(2012) LGD and EL Estimation by using multistage model based on observed debt-collection data, *FSA Institute Discussion Paper Series*, DP 2012-6(In Japanese).

Moody's investors service(2011) Corporate Default and Recovery Rates 1920-2010, *SPECIAL COMMENT,* JUNE 2011.

Ogi, K., Toshiro,M. and Hibiki, N.(2014) Effect of Firm Age in the Credit Scoring Model for Small Sized Firms, *Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference 2014 Abstracts*,27.

Ogi, K., Toshiro,M. and Hibiki, N.(2015) Empirical Study of Recovery Rate Model for Small Sized Firms, *JAFEE journal*, 168-208(in Japanese).

Witzany, J. (2011)A Two Factor Model for PD and LGD Correlation, *Bulletin of the Czech Econometric Society*, 18.