Improving the Recovery Rate of Unsecured Debt in Multistep Workout Processes

Johannes Kriebel,[1*] Kevin Yam[2]

[1]University of Münster
[2]Seghorn AG

*E-mail: johannes.kriebel@wiwi.uni-muenster.de.

January 20, 2019

Abstract

Forecasting recovery rates is usually mainly concerned with the prediction of a customer's capacity to repay defaulted debt given a set of contract, demographic and macroeconomic information. In this paper, we suggest a model to determine workout strategies given the stylized fact that debtors react differently to weaker and stronger workout actions (e.g. mail contact, phone contact, legal proceedings). Debtors are ceteris paribus more likely to pay when faced with stronger workout actions but stronger workout actions tend to be more costly for the creditor. There is therefore a need to find optimal workout strategies for the creditor. We suggest a regression model in which the creditworthiness of a debtor is a linear combination of debtor characteristics. The reaction towards workout actions is incorporated by using skewed error terms that make payments more likely for strong workout measures and less likely for weak workout measures. We outline, how the model coefficients and skewness parameters can be estimated from empirical data. Using simulated data, we show that model predictions can be substantially improved using our model to account for the debtor's behaviour when compared to predictions from a logistic regression model. We further show that the workout success in terms of the sum of net payments can be improved substantially by using the suggested model to choose appropriate workout actions when compared with other alternative ways to determine the workout strategy.

We thank the participants of the CRCC 2017 in Edinburgh, the participants of OR2017 in Berlin and the members of the Finance Center Muenster for useful feedback that helped to improve this work.

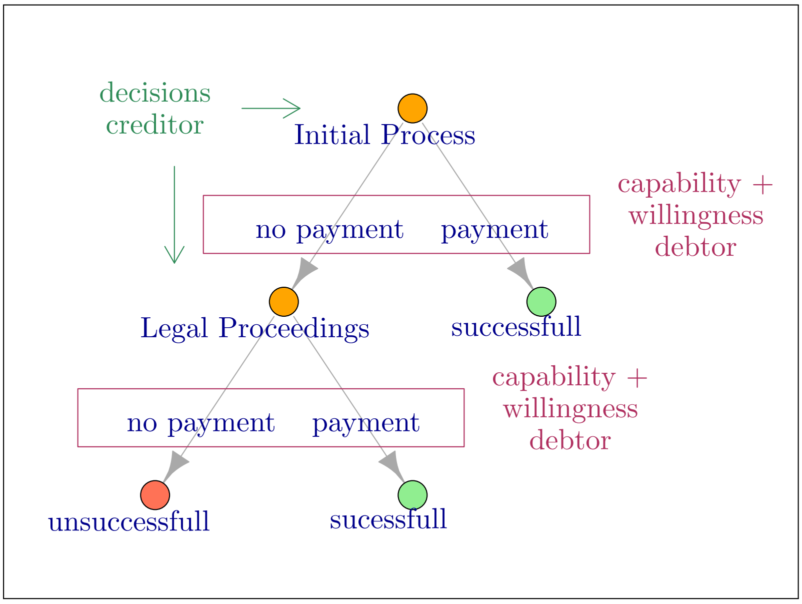
1 Introduction

Over the past two decades the credit risk literature has gathered many insights into various aspects of the loss given default (LGD) and the recovery rate (RR) on bank loans. The LGD/RR on bonds have attracted attention earlier but have been in the focus of many publications of the past two decades as well. One aspect that is mentioned in many places in the literature (e.g. Davydenko and Franks, 2008, Han and Jang, 2013, Ingermann et al., 2016, Khieu et al., 2012, and Qi and Yang, 2009) but is rarely assessed in more detail is the choice of workout actions.

We use Figure 1 to illustrate how an important aspect of the the LGD/RR is neglected this way. When making predictions of the LGD/RR, publications are mostly concerned with the capability of a debtor to repay debt and do not consider that the debtor might be unwilling to repay (Matuszyk et al., 2010). Whether a payment is received could therefore be argued to depend on the willingness to pay besides the sole capacity. This willingness is further likely to depend on the severity of workout actions. Practically, a difficult debtor might be more likely to repay when faced with legal actions compared to simple written reminders. Ceteris paribus, the creditor aims at choosing the most effective workout actions. As a consequence, the LGD/RR is also dependent on the decision of the creditor.

Figure 1: Two-stage collection process

This figure displays the exemplary structure of a two-stage collection process.



The available workout actions are associated with different costs (de Almeida Filho et al., 2010). The costs could be argued to increase with the severity of the workout action. Telephone calls are more costly than standardized letters, legal actions are more costly than telephone calls. Hence, besides choosing the most effective workout actions, there is therefore a need to determine efficient workout actions making the choice of workout actions an optimization problem. Dealing with this optimization problem has rarely been addressed in the literature (e.g. Makuch et al., 1992, de Almeida Filho et al., 2010, Matuszyk et al., 2010, and Thomas et al., 2016).

Our paper contributes to the small literature on choosing workout actions by suggesting a novel model that makes predictions of the payment probability in individual steps in a two-stage workout process as given in Figure 1 by explicitly modeling the severity of workout steps. This is done by allowing the error terms of a binomial regression model to vary with the workout action. The resulting predicted payment probabilities for the individual steps can be used to decide on the actions conducted given the associated costs.

The benefits of the model are illustrated using simulated collection data. Our results show that in a situation that applies to our model assumptions, this enables better predictions of the recovery rate compared to predictions from the logistic regression. We further show that the net recoveries, as the recoveries deducted by the incurred costs, can be considerably increased using our model.

The remaining part of the paper proceeds as follows: In the second section, we present an overview of the literature that mentions the workout process as an influence on LGD/RR and the literature that addresses the problem of choosing workout actions. In the third section, we present the prediction model. In the fourth section, we outline the results of the simulation study considering the prediction accuracy of the model and the performance in choosing workout actions. Section five concludes.

2 Literature Review

2.1 Workout Actions as an Influence on the LGD/RR

There is a comprehensive literature on many aspects of the LGD/RR. Information on workout actions is recognized to influence the LGD/RR by many authors and is commonly included in prediction models. These include Davydenko and Franks (2008), Han and Jang (2013), Ingermann et al. (2016), Khieu et al. (2012), and Qi and Yang (2009) but are not limited to these examples. Given this fact, it is surprising to notice that workout process information is often only a side-note even in these publications and optimizing the workout process or determining good workout actions have rarely been discussed in the LGD/RR literature.

Among the studies including workout process information, Davydenko and Franks (2008) study the influence of bankruptcy laws in a cross-country study of France, Germany, and the UK. They find that bankruptcy laws have an influence on differences of the RR in the three countries. This holds true despite their finding that banks adapt their lending conditions in order to cope with the different levels of riskiness.

The influence of workout actions is explicitly discussed in Han and Jang (2013). The authors include dummy variables for workout decisions in a LGD prediction model on defaulted Korean SME loans. They find that a model only including workout characteristics is able to explain a larger part of the LGD variation than a model only including firm specific characteristics. Han and Jang (2013) suggest to use this model to support decisions on workout actions.

Ingermann et al. (2016) include a dummy variable for amicable agreements in a prediction model of the LGD in German SME loans and individual loans. They find that arriving at an amicable agreement is positively related to the RR. Khieu et al. (2012) include information of bankruptcy types in a model for bank loans from Moody's Ultimate Recovery Database. They find the bankruptcy type to influence the RR. Qi and Yang (2009) study a data set of mortgage loans from the United States and include information of the foreclosure process and the time of property sale in a prediction model. They find these characteristics to influence the LGD.

2.2 Determining Workout Actions as an Optimization Problem

Although the choice of workout actions has rarely been addressed as an optimization problem, there are some previous results on this issue. These include Matuszyk et al. (2010), de Almeida Filho et al. (2010), Makuch et al. (1992), and Thomas et al. (2016).

Matuszyk et al. (2010) model the workout process as a decision tree where each node of the tree is one workout action. To determine the loss given default for a node, respectively a workout action, they suggest a two-stage regression method (modeling full payments, $LGD = 0$, and partial payments, $LGD > 0$, separately). Given the predicted outcome, a bank could decide whether a workout action is profitable. The authors illustrate their model using a data set of loans provided by a financial institution from the UK.

de Almeida Filho et al. (2010) model a linear workout process with workout steps in increasing severity. A collection department starts with the weakest workout action and proceeds with increasingly strong workout actions until the process is either successful or terminated by the collector. The recovery as a fraction of the exposure is a function of time with decreasing marginal recoveries for each workout action. The individual workout actions have specific costs per time period. The model aims at determining the best points in time to switch from one workout action to the next. de Almeida Filho et al. (2010) use a set of European bank loans to illustrate how the repayment functions can be estimated and the transition times to stronger workout actions are determined.

Makuch et al. (1992) build an optimization model by estimating transition matrices for defaulted consumer credit at GE Capital. The matrices state the probability of a delinquent debtor to migrate from one delinquency state (number of months in arrears) to another delinquency state. These transition matrices are estimated conditional on the balance range, the debtor rating and the workout action employed. Given the debtor characteristics and a sequence of workout actions, the probability of transitioning from a delinquent to a recovered state can be calculated and an optimal workout strategy can be derived.

Thomas et al. (2016) notice that delinquent borrowers tend to jump between paying and non-paying time periods. They suggest methods to choose the wright-off policies for a contract. A debtor might for example pay three months in a row and then skip two subsequent months and so forth. Thomas et al. (2016) aim at modeling the probability of transitioning from paying to non-paying states using two approaches. The first approach is a markov chain where the probability of transitioning to a paying or non-paying state is only dependent on the current state of a debtor. The process continues until the debt is either fully repaid or the claim is written off. The second approach is a hazard rate model of how long the debtor remains in a paying or non-paying sequence. The two approaches can be used to estimate the LGD for different write-off policies. The models are illustrated using a data set of personal loans from the UK.

3 Model

3.1 Modelling the Collection Process

In Section 2, we outlined that the choice of workout actions is recognized to influence the LGD/RR in the literature. We further presented four approaches to determine appropriate workout actions.

These four methods build on different sets of collection process information. In the approaches of de Almeida Filho et al. (2010) and Thomas et al. (2016) very detailed information of payments in a time period (e.g. week or month) are required. de Almeida Filho et al. (2010) further need very detailed data of the workout actions applied and the related cost for each time period. The approach of Makuch et al. (1992) further includes information on the status of delay. Makuch et al. (1992) need very large data sets to estimate the transition matrices for many combinations of customer quality, status of delay and workout actions in particular. This might be problematic when workout actions are rarely applied to certain customer groups.

Our model is most closely related to Matuszyk et al. (2010). Matuszyk et al. (2010) suggest to model the outcome of each workout action with an individual regression model. We further build on the idea of modeling the workout process in analogy to potential barriers in quantum mechanics[1]. A particle encountering a potential barrier has a probability of transmitting this barrier. This probability is dependent on the energy of the particle and the height of the potential barrier. In a similar way, we model the probability of workout actions being successful as dependent on the credit quality of the debtor and the strength of the workout action. In the potential barrier model, the energy of a particle is constant irrespective of whether a potential barrier was transmitted or not. In a similar way, we model the creditworthiness of a debtor to be constant over the workout process. The success of sequential workout actions is therefore assumed to be independent from each other and only depends on the constant debtor characteristics and the varying levels of the workout action strength.

In order to model the workout action strength, we suggest a binary regression model that allows the distribution of an error term ε_{iw} for each debtor i to vary with the severity of the workout action w. This way we receive a regression model that makes a prediction of the probability of receiving a payment in an individual workout step. The overall probability of payment can be derived from these individual probabilities as the payment probabilities in individual workout steps are assumed to be independent.

3.2 Modelling the Severity of Workout Actions

Single step workout process

Binary outcome regression models such as the logistic regression can be described as a latent variable model. For the logistic regression, there is an observable variable $y_i \in \{0, 1\}$ that is equal to one when the latent variable $y_i^* = x_i\beta + \epsilon_i$ is larger than zero and equal

[1] Refer to Griffiths (2017) for an introduction to basic principles of quantum mechanics.

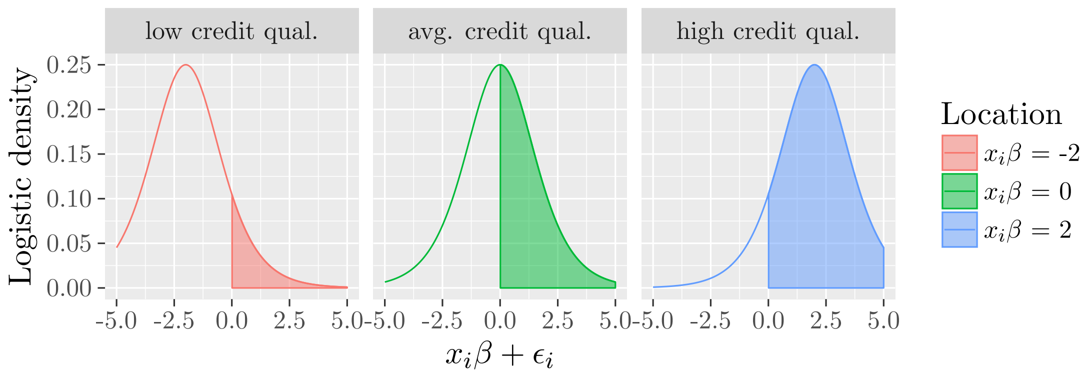
to zero otherwise (Equation 3.1). The first case, $y_i = 1$, is considered here to mark a sufficient payment as an outcome and $y_i = 0$ mark an insufficient payment. The latent variable is given by a linear combination of the debtor characteristics x_i with the regression coefficients β and an error term ϵ_i. For a given level of the debtor characteristics x_i and regression coefficients β, the probability of a payment $P_S(Y_i = 1|X)$ can be derived using the density of the logistic distribution (refer to Equation 3.2 and to Figure 2 for a graphical illustration). $F_L(0; x_i\beta, 1)$ is the value of the logistic cumulative density function at zero with a location of $x_i\beta$ and a variance parameter (scale) of one.

$$y_i = \begin{cases} 1 & x_i\beta + \epsilon_i > 0 \\ 0 & x_i\beta + \epsilon_i \leq 0 \end{cases} \tag{3.1}$$

$$P_s(Y_i = 1|X) = 1 - F_L(0; x_i\beta, 1) = \frac{exp(x_i\beta)}{1 + exp(x_i\beta)} \tag{3.2}$$

Figure 2: Density of the logistic distribution for different locations

This figure displays the density of the logistic distribution for the location parameters -2, 0, and 2, respectively, and a scale parameter equal to one. The filled area is the area under the curve for $x_i\beta+\epsilon_i > 0$. This area is the probability of receiving a full payment.



Multistep workout process

Our aim is to derive a model that assigns payment probabilities to individual workout steps. We build on the latent variable model but use a different distribution of the error terms ε_{iw} for the workout steps $w \in 1, 2$. This way, the latent variable model is written as:

$$y_{iw} = \begin{cases} 1 & x_i\beta + \epsilon_{iw} > 0 \\ 0 & x_i\beta + \epsilon_{iw} \leq 0 \end{cases} \tag{3.3}$$

The distribution parameter that we alter is the skewness of the error terms. In order to have a different skewness parameter for the error terms, we replace the logistic distribution with the skew normal distribution as described in Azzalini (1985). This family of distributions has the normal distribution as a special case. However, by varying a an additional skewness parameter α, besides a location and scale parameter, the distribution can obtain various levels of right and left skewness, resulting in higher or lower expected payments. In the following, we use α_1 to be the skewness parameter for the weak workout step and α_2 to be the skewness of the strong workout step. The probability of a step being successful can be derived from the equations 3.4 and 3.5 where T is a function defined in Owen (1956) and Φ is the cumulative normal density function.

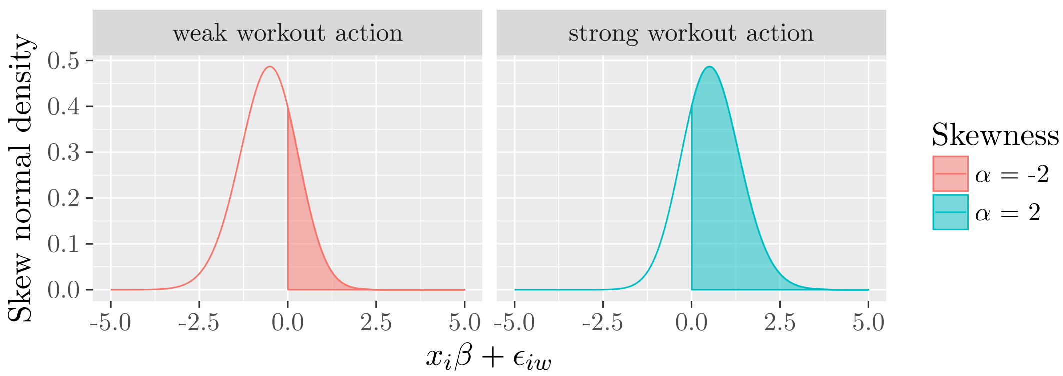
$$P_m(Y_{iw} = 1|X) = 1 - F_{SN}(0; x_i\beta, 1, \alpha_w) \tag{3.4}$$

$$F_{SN}(0; x_i\beta, 1, \alpha) = \Phi(-x_i\beta) - 2T(-x_i\beta; \alpha) \tag{3.5}$$

The density of the skew normal distribution is illustrated in Figure 3 for the skewness parameters -2 (left panel) and 2 (right panel) with the location of zero and the scale parameter one. The distribution on the left panel is right skewed and is considered to be the distribution of the latent variable in a weak workout step in terms of our model. The distribution on the right belongs to a stronger workout step in terms of our model. Given that the regression parameters β and the skewness parameters α_1, α_2 are known,

Figure 3: Density of the skew normal distribution for different skewness parameters

This figure displays the density of the skew normal distribution for the skewness parameters -2 and 2, a scale parameter equal to one and a location parameter equal to zero. The filled area is the area under the curve for $x_i\beta + \epsilon_i > 0$. This area is the probability of receiving a full payment.



the probability of a workout process with the workout strategy $W \in \{(0), (1), (1, 2)\}$ being successful $P_M(Y_i = 1|X, W)$ is given by the equations 3.6, 3.7 and 3.8.

$$P_M(Y_i = 1|X = x_i, W = (0)) = 0 \tag{3.6}$$

$$P_M(Y_i = 1|X = x_i, W = (1)) = P_m(Y_{i1} = 1|X = x_i) \tag{3.7}$$

$$\begin{aligned} P_M(Y_i = 1|X = x_i, W = (1,2)) &= P(Y_{i1} = 1|X = x_i) \\ &+(1 - P(Y_{i1} = 1|X = x_i))P(Y_{i2} = 1|X = x_i) \end{aligned} \tag{3.8}$$

3.3 Parameter Estimation

Considering the model we present in Section 3.2, there is the question of how to estimate the linear parameters β and the additional skewness parameters α of the model from LGD/RR data. We suggest two estimation approaches, one minimizing the root mean squared errors (RMSE) and another one minimizing the mean absolute errors (MAE).

More specifically, we proceed by first calculating the expected recovery rate for each exposure given the workout strategy w_i applied to the debtor i for a set of parameters β and α. This can be derived from the equations 3.6, 3.7 and 3.8. In a second step, we calculate the MAE and RMSE for each of these expectations compared to the actual payment in the workout process. N is the total number of debtors.

$$E(RR_i|X = x_i, W = w_i) = P_M(Y_i = 1|X = x_i, W = w_i) \tag{3.9}$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N} |E(RR_n|X = x_n, W = w_n) - Y_i|) \tag{3.10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(E(RR_n|X = x_n, W = w_n) - Y_i)^2} \tag{3.11}$$

We determine the model parameters by minimizing the MAE and the RMSE over the values of β and α. The solutions for the parameters are derived using the Nelder-Mead method as in Nelder and Mead (1965).

4 Simulation Study

4.1 Data Generation

The main purpose of our model is to support the decision making on workout actions in defaulted debt contracts. In order to illustrate its benefits, we generate a simulated data set that replicates (simplified) collection department data and applies to a situation with two workout steps that have a different severity and different incurred costs. The key

Figure 4: Simulated data set

This figure displays the key characteristics of the data set in the simulation study.

Number of debtors:
$N = 100,000$

Debtor characteristics:
Credit score $\sim$ Uniform distribution on interval $[300, 600]$
Gender $\sim$ Bernoulli distribution with $p = 0.5$
Insolvent $\sim$ Bernoulli distribution with $p = 0.1$

Coefficients:
$\beta = (30, -0.1, 10, -20)$

Workout step costs:
$c_1 = 0.05$, $c_2 = 0.15$

Skewness:
$\alpha_1 = 0$, $\alpha_2 = 4$

Workout strategy:
$W_i = (1, 2)$ for all debtors

characteristics of the model are displayed in Figure 4.
We model data for $100,000$ hypothetical debtors. These debtors have three characteristics. The credit score follows a uniform distribution with possible realizations between 300 and 600. The gender follows a Bernoulli distribution with a probability of 0.5. The insolvency follows a Bernoulli distribution with a probability of 0.1. The regression coefficients β indicate a negative relation of the credit score, a positive relation of the gender dummy and a negative relation of the insolvency status to the payment probability.

There are two workout steps in the process. The costs of conducting the individual workout steps are given by c_1 and c_2. The second workout step is more costly compared to the first step. In line with our model, the second workout step has a higher α, indicating a higher payment probability compared to the first step.

In order to derive payment information from the data, we first use our model to calculate payment probabilities for each of the two workout steps. Afterwards, we use these payment probabilities to generate Bernoulli random numbers indicating whether a debtor is willing and able to pay in a specific workout step. Using the workout strategy applied, we determine, whether this leads to a workout process that is successful. For the sake of simplicity, we assume that both workout steps are conducted in all cases. When the debtor is willing and able to pay in at least one process step, the process is considered to be successful. When the first step is successful, there are no costs resulting from the second workout step.

4.2 Prediction Accuracy

We use the simulated data in order to compare the performance of the new prediction model with the performance of the logistic regression as a typical prediction model in LGD/RR. The results are compared according to three goodness-of-fit measures: The misclassification rate[2], the MAE and the RMSE. We compare the results in-sample and out-of-sample. The in-sample assessment is conducted by building the models on the whole $100,000$ observations and calculating the error measures on the same set of debtors. In the out-of-sample assessment, we fit the models on the first $10,000$ debtors and calculate the error measures on the remaining $90,000$.

Table 1: Goodness of fit - In-sample

This table displays the in-sample goodness of results for the skew normal regression fitted by minimizing the MAE and the RMSE. The results for the skew normal regression are compared to the logistic regression results using a Welch test.

Measure	Log. Regr.	Skew. Norm. (min. MAE)	Skew. Norm. (min. RMSE)
Miscl. Rate	0.1759	0.1650^{***}	0.1676^{***}
MAE	0.3323	0.1922^{***}	0.2139^{***}
RMSE	0.4015	0.3976^{**}	0.3900^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

The in-sample results are presented in Table 1. The results for the logistic regression are presented in the first column and the results for the skew normal binomial regressions are listed in the second and third column. The new prediction method excels considering all three error measures. About $1,000$ more cases are correctly classified compared to the logistic regression. The the MAE is considerably lower and the RMSE is slightly lower compared to the logistic regression. We test the difference to the logistic regression more formally using a Welch test. All differences are significant at the 1% or at least at the 5% level in one case.

The out-of-sample results are presented in Table 2. These results are clearly in line with the in-sample results. The new regression model excels in all three error measures. The

[2] The misclassification rate is given by the proportion of wrongly classified cases.

Table 2: Goodness of fit - Out-of-sample
This table displays the out-of-sample goodness of results for the skew normal regression fitted by minimizing the MAE and the RMSE. The results for the skew normal regression are compared to the logistic regression results using a Welch test.

Measure	Log. Regr.	Skew. Norm. (min. MAE)	Skew. Norm. (min. RMSE)
Miscl. Rate	0.1762	0.1648^{***}	0.1673^{***}
MAE	0.3313	0.1922^{***}	0.2138^{***}
RMSE	0.4014	0.3977^{**}	0.3900^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

differences to the logistic regression are significant at the 1% level except one case, where the difference is significant at the 5% level.

4.3 Determining Workout Strategies

In this step of the analysis, we discuss how the new model can be employed to make workout decisions. We compare four competing methods (see Figure 5):

(a) Conduct both workout actions in all cases.

(b) Choose between no, one or two workout action randomly.

(c) Conduct all workout actions that are profitable in logistic regression.

(d) Conduct all workout actions that are profitable in binomial skew normal regression. (Action two can only be conducted after action one.)

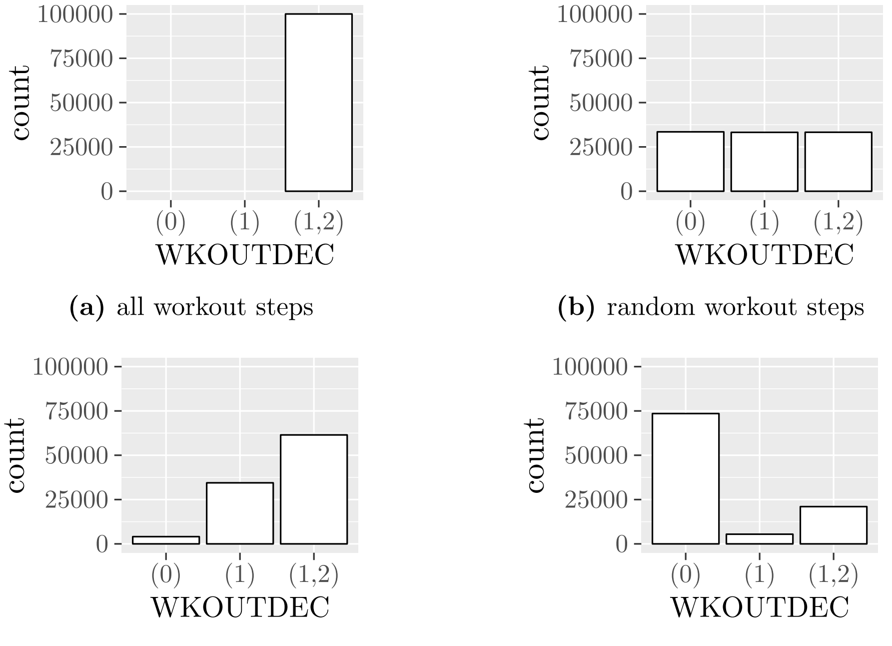
The frequencies of the resulting workout decisions for the four workout strategies are plotted in Figure 5. In the case (a), all debtors are treated with both workout actions. In the randomly selected workout action (b), debtors are randomly assigned to three groups that are treated with no, one or two workout actions. When determining the workout decision using the logistic regression (c), we first calculate the expected payment and all workout actions are conducted for which the expected recovery is above the cost of the workout actions. The same is done for the new regression model (d).

The resulting net recoveries are listed in Table 3 for the data parameters as outlined in Section 4.1 and the model parameters as fit in Section 4.2. Given this configuration, randomly choosing the workout strategy performs the worst. Choosing the workout strategy by logistic regression is superior to the random selection but still inferior to conducting the full two steps in all cases. Choosing the workout strategy according to the new regression model is superior to all three alternatives.

In Table 3, we further list the recoveries and costs incurred by individual workout actions

Figure 5: Workout strategy decisions

This figure displays the frequency of workout decisions conditional on four different workout strategies.



(a) all workout steps

(b) random workout steps

(c) logistic regression

(d) bin. skew normal regression

in more detail. One noteworthy aspect is that the methods (a), (c) and (d) all recover a similar amount in step one. There is, however, a large difference between these three methods in the costs incurred in the first step. The costs are on a similar level for method (a) and (c) but on a far lower level for method (d). The recoveries from the second step are very different for the first three methods compared to the fourth method that recovers much less, which is in line with the fact the much less cases are worked out (refer to Figure 5). The workout cost are much lower in method (d) accordingly. Judging from Figure 5, the central difference between method (d) and the other three methods is that it focuses on a small number of profitable cases and rather lets go some additional recoveries for the sake of saving workout costs.

The difference between the predictions from the logistic regression and the new regression model is more visible in a different way in Figure 6. What is apparent from the upper panel is that the payment probability for the whole process is much more accentuated with the size of the linear predictor $x_i\beta$ compared to logistic regression. Logistic regression produces much more intermediate predictions while the true model and the fitted binomial skew normal regression have much more accentuated payment probabilities. This is even

Table 3: Net recoveries for different workout strategies

This table displays the net recoveries for the four workout strategies in the first column. More detailed information on the recoveries and the costs of the individual workout steps are listed in the columns two to five.

Strategy	net recovery	recovery w_1	cost w_1	recovery w_2	cost w_2
(a) all	12637.45	15583.00	5000.00	14717.00	12662.55
(b) random	7729.70	10383.00	3335.75	4906.00	4223.55
(c) logistic regr.	11824.70	15583.00	4795.80	7921.00	6883.50
(d) bin. skew. norm.	13843.90	14997.00	1323.95	1074.00	903.15

clearer in the second panel that displays the predictions of the two models and the actual payment probabilities in ascending order. One can infer from this plot that the new model is much more decisive of classifying debtors to be payers or non-payers while the logistic regression has much more even predictions. This way, many debtors that the new model clearly considers to be non-payers and are therefore not exposed to workout actions become eligible for entering the first or second workout step using the logistic regression despite the cost of these workout steps. The main benefit in the prediction accuracy and in choosing workout actions appears to be that the new model is more able to identify good from bad debtors.

5 Conclusion

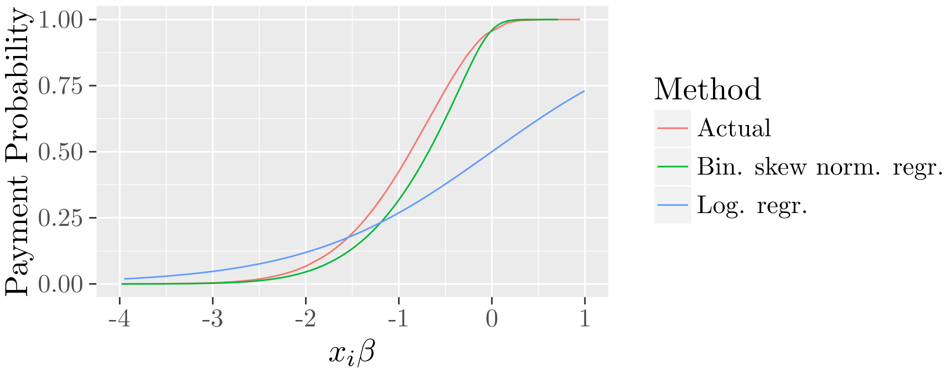
The influence of workout process actions is mentioned in many contributions to the LGD/RR literature. However, very few publications address methods of determining appropriate workout actions.

We address this issue by suggesting a novel prediction model that explicitly models the severity of workout process actions and the reaction of the debtor towards these actions. This is done by allowing the skewness of error terms in a binomial regression model to vary with the severity of the workout actions. This way, predictions of the payment probability in individual workout steps can be derived. These probabilities can be employed to choose profitable workout actions.

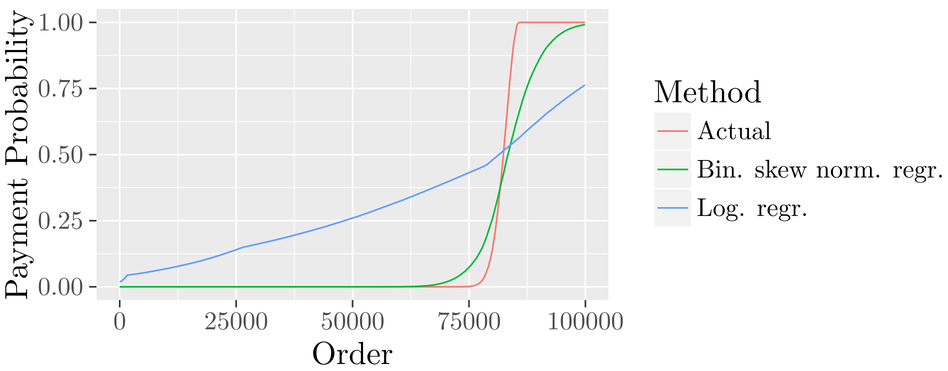
Besides adding a novel approach to a small strand of literature, our model requires less detailed information on monthly payments, conducted workout actions and incurred costs compared to de Almeida Filho et al. (2010) and Thomas et al. (2016). Compared to Makuch et al. (1992) who estimate transition matrices for many different cases of arrear, workout actions and debtor quality, there is less data needed. Compared to Matuszyk et al. (2010) our approach is estimates a model for multiple workout steps simultaneously.

Figure 6: Linear term and predicted payment probabilities

This figure displays the predicted payment probability for the logistic regression and the binomial skew normal regression plotted over (a) the linear part of the regression models $x_i\beta$ and (b) an ordered index. The actual payment probability from the simulation is further added to the plot.



(a) payment probability plotted over linear predictor



(b) ordered payment probability values

We assess the performance of our model in a simulation study. In a situation that applies to our model assumptions, it can substantially improve the quality of predictions compared to logistic regression. We further show that this enables better workout decisions that can considerably increase the net recoveries.

Further work could address whether the assumption that the workout action success is independent is justified. In addition, it could improve the optimization methods used in this work. The current approach is limited to a two-stage process. Further work could extend this to longer and more complex process steps. Additional work could further handle partial payments in workout steps.

References and Notes

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*(2), 171–178.

Davydenko, S. A. and J. R. Franks (2008). Do bankruptcy codes matter? A study of defaults in France, Germany, and the U.K. *The Journal of Finance 63*(2), 565–608.

de Almeida Filho, A. T., C. Mues, and L. C. Thomas (2010). Optimizing the collections process in consumer credit. *Production and Operations Management 19*(6), 698–708.

Griffiths, D. J. (2017). *Introduction to quantum mechanics* (Second edition ed.). Cambridge: Cambridge University Press.

Han, C. and Y. Jang (2013). Effects of debt collection practices on loss given default. *Journal of Banking & Finance 37*(1), 21–31.

Ingermann, P.-H., F. Hesse, C. Bélorgey, and A. Pfingsten (2016). The recovery rate for retail and commercial customers in Germany: A look at collateral and its adjusted market values. *Business Research 9*(2), 179–228.

Khieu, H. D., D. J. Mullineaux, and H.-C. Yi (2012). The determinants of bank loan recovery rates. *Journal of Banking & Finance 36*(4), 923–933.

Makuch, W. M., J. L. Dodge, J. G. Ecker, D. C. Granfors, and G. J. Hahn (1992). Managing consumer credit delinquency in the US economy: A multi-billion dollar management science application. *Interfaces 22*(1), 90–109.

Matuszyk, A., C. Mues, and L. C. Thomas (2010). Modelling LGD for unsecured personal loans: Decision tree approach. *Journal of the Operational Research Society 61*(3), 393–398.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer Journal 7*(4), 308–313.

Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics 27*(4), 1075–1090.

Qi, M. and X. Yang (2009). Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Finance 33*(5), 788–799.

Thomas, L. C., A. Matuszyk, M. C. So, C. Mues, and A. Moore (2016). Modelling repayment patterns in the collections process for unsecured consumer debt: A case study. *European Journal of Operational Research 249*(2), 476–486.