ESTIMATING RECOVERY RATE AND TIME TO LIQUIDATE FOR NPLs

Roberto Rocci

Alessandra Carleo

Maria Sole Staffa

ISSN 2611-9633

Working Papers (Dipartimento di Economia Aziendale)

[online]

Working Paper Numero 16, 2021

Collana del Dipartimento di Economia Aziendale

I Working Paper del Dipartimento di Economia Aziendale svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali.
La loro pubblicazione è soggetta all'approvazione del Comitato Scientifico.
Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall'art. l del D.L.L. 31 agosto 1945 n. 660 e successive modifiche.
Copie della presente pubblicazione possono essere richieste alla Redazione.
Esemplare fuori commercio ai sensi della Legge 14 aprile 2004 n. 106.

REDAZIONE
Dipartimento di Economia Aziendale
Università degli Studi Roma Tre
Via Silvio D'Amico, 77
00145 Roma – Italia
Email:
ricerca.economiaaziendale@uniroma3.it

Abstract

The objective of the present paper is to propose a new method to measure the recovery performance of a portfolio of non-performing loans (NPLs) in terms of recovery rate and time to liquidate. The fundamental idea is to draw a curve representing the recovery rates during time, here assumed discretized, for example, in years. In this way, the user can get simultaneously information about recovery rate and time to liquidate of the portfolio. In particular, it is discussed how to estimate such a curve in presence of right censored data, i.e. when the NPLs composing the portfolio have been observed in different time periods. Uncertainty about the estimates is depicted trough confidence bands obtained by using the non-parametric Bootstrap. The effectiveness of the proposals is shown by applying the method to a real financial data set about some portfolios of Italian unsecured NPLs taken in charge by a specialized operator.

Keywords: Recovery rate, NPLs, Censored data.

J.E.L. Classification: C13, C24, C41, G20, G21.

Corresponding author: Alessandra Carleo, email address: alessandra.carleo@uniroma3.it

SUMMARY –1 Introduction. 2 Recovery rate and time to liquidate of a portfolio. 3 Estimating the recovery rate curve from censored data. 4 Application. 5 Conclusions and final remarks.

1. Introduction

Non Performing Loans (NPLs) are loans whose collection by banks is uncertain. They constitute one of the three Non Performing Exposure (NPE) categories, which are distinguished because of the different probability of credit recovery at maturity:

- Past Due are exposures that exceed the credit limits of more than 90 days;
- Unlikely to Pay (UTP) are exposures for which the bank considers it unlikely that the borrower will fully fulfil its contractual obligations without recourse to actions such as the enforcement of the guarantee;
- Non Performing Loans (NPL) are exposures in state of insolvency.

As Resti and Sironi (2007) point out, an effective recovery depends on the characteristics of the exposure (presence of collateral, degree of effectiveness), of the counterparty (e.g. industry, country and legal framework), on macroeconomic factors like the state of the economy, and on internal factors like the efficiency of a bank in recovering its money, for instance in dealing with out-of-Court settlements.
There is a NPL market that offers banks the opportunity to get rid of non-performing loans by selling them to specialized operators who deal with recovery.
The main method for determining the value of Non Performing Loans is that of discounted financial flows, according to which the value of the loans is equal to the sum of the expected income flows, discounted at a rate consistent with the expected unlevered return of the investor and net of the related recovery costs.

In the case of a performing loan, the borrower is expected to pay principal and interest at the agreed deadlines with a high level of probability (one minus the probability of default, generally low). In this case, the uncertainty in the valuation is limited to the determination of the discount rate to take into account the general market circumstances of the rates and the specific risk of the debtor.
In the case of Non Performing Loans, the uncertainty concerns not only the discount rate but also the amount that will be returned and the time of return. In fact, the probability of default is now equal to one, in the case in which the transition to non-performing has already occurred, or is in any case very high, if the credit is in the other categories of impaired loans (unlikely to pay).

The valuation methodologies currently used on the market are therefore based primarily on forecast models of the amount of net repayments

expected from receivables and related collection times. The operation is not trivial and is carried out with different forecast models. The choice of the forecast model of expected net flows essentially depends on the type of credit and on the information available to the evaluator.
To choose the flows forecast model, it is first necessary to consider whether a real guarantee (typically a mortgage or pledge) on an existing asset with a market value covers the credit. In this case, the flow forecast model is based on the lesser of the value of the asset covered by the guarantee, the amount of the guarantee and the value of the credit, and on the timing for its judicial sale. The valuation methods to apply are those based mainly on the compulsory recovery of the credit, however providing for the possibility of recovering the credit through an out-of-court agreement in some cases.
Forecast models are generally based on:

a) information available to the creditor;
b) public information;
c) information acquired and processed as part of the analysis.

The availability of one information rather than another, radically changes the articulation and the degree of detail that the evaluator can give to the flows forecasting models and consequently to the evaluation methods.
The forecast model with the aforementioned limits uses all the relevant information available to determine the estimated flows and related timing.
The forecasting models can be traced back to three types, which can be partially combined with each other: models based on judicial recovery; models based on the debtor's restitution capacity; statistical forecasting models.

The estimation methodology for recovery rate, which we are interested in for $NPLs$, was faced in the more general context of Basel II.

In Basel II, the Basel Committee in 2001 proposed an internal ratings-based (IRB) approach to determine capital requirements for credit risk. This IRB approach grants banks permission to use their own risk models, or assessments, to calculate regulatory capital. Under the IRB approach, banks are required to estimate the following risk components: probability of default (PD), loss given default (LGD), exposure at default (EAD) and maturity (M). Since Basel II's capital requirement calculation depends heavily on LGD, financial institutions have put more emphasis on modelling such a quantity in recent years.
Unlike the estimation of PD, which is well established, LGD is not so well understood and still subject to research. Improving LGD modelling and estimation can help financial institutions to assess their risk and regulatory capital requirement more precisely, as well as improving debt management.
LGD is defined as the proportion of money financial institutions fail to gather during the collection period, given the borrower has already defaulted. Conversely, Recovery Rate (RR) is defined as the proportion of money financial institutions successfully collected minus the administration fees during the collection period, given the borrower has already defaulted.
That means LGD = 1- RR.

Recovery rate (or LGD) can be estimated using both parametric and non-parametric methods. Mainly, recovery rate is estimated using parametric methods and considering a one-year time horizon.
Methods used in literature, among others, are: classical linear regression, regularized regression like Lasso, Ridge, Elastic-net, etc. (Hastie et al., 2009), beta regression, inflated beta regression, two-stage model combining beta mixture model with a logistic regression model (Ye and Bellotti 2019).

In the case of NPLs, in our opinion, in investigating the recovery process of defaulted exposures the focus must be not only on the recovered amounts, but also on the duration of the recovery process, the so-called time to liquidate (TTL).

Cheng and Cirillo (2018) propose a model that is able to learn, improving its performances over time, using the mechanism of Bayesian update (in a machine learning context). The final goal is to predict the possible recovery trajectory of a counterpart, not only on the basis of the available data, but also with the possibility of using experts' judgements and other a priori knowledge to mitigate historical bias, at least partially.

Our purpose is to introduce a particular method to measure the performance of a NPLs portfolio in terms of recovery rate (RR) and time to liquidate (TTL). The idea is to represent the recovery process as a curve showing how the RR is distributed during the time. We will also propose a method to estimate such a curve when some data are censored. The plan of the paper is the following. In section 2, we show how the recovery curve is defined. The method of estimation in case of censored data is discussed in section 3. In section 4, the effectiveness of the proposal is shown through an application on real data, while some conclusions and final remarks are discussed in section 5.

2. Recovery rate and time to liquidate of a portfolio

Given a single NPL it is clear how to define and compute its RR and TTL. The extension of such definitions to the case of a NPLs portfolio is not trivial because the two quantities are strictly related. To make clear what do we mean by "recovery rate" and "time to liquidate" and how they are related in the case of a portfolio, we have to answer to several questions about the RR. For example: when do we measure the RR? When the last NPL in the portfolio has been liquidated or after a given period? And, in the latter case, how do we choose the length of the period? Similarly for the TTL: when do we measure the TTL? When the last NPL has been liquidated or when a significant part of the portfolio has been recovered? And, in the latter case, how much is the significant part?

The aforementioned questions make clear that the measurement of the RR cannot disregard the measurement of the TTL and vice versa. How do we face with this problem?
First, we note that to measure the TTL when the last NPL has been liquidated could lead to measures highly affected, and biased, by anomalous NPLs with long TTLs and small EAD. It follows that the TTL should be measured when the RR becomes significant. It remains to understand what is "significant". Second, in many cases the user needs a more complete information rather than only two numbers: RR and TTL. It would be better to know how the RR increases during the time. This would also help in choosing at what RR point to measure the TTL. For the aforementioned reasons, we decide to measure the behaviour of the RR during the time through what we called the "recovery curve". Such a curve is built in the following way.

Let us consider a portfolio of K NPLs. For each of the K NPLs the debt exposure at default is EAD_k (exposure at default of the k-th NPL) and the total exposure for the bank is $EAD = \sum_{k=1}^{K} EAD_k$. Assume I time intervals (of the delay of payment) from the default, in time t_0, to time t_I, i.e. the valuation date. Let $p_{k,i}$ be the recovery of the k-th NPL, in the i-th interval (of delay), i.e. $(t_{i-1}, t_i]$, with $k \in \{1, 2, \ldots, K\}$ and $i \in \{1, 2, \ldots, I\}$. The portfolio recovery in time interval i equals $p_i = \sum_{k=1}^{K} p_{k,i}$, that is the total recovery, for all the K debt positions, in the i-th time interval of delay. Consequently, after i time intervals of delay, i.e. by the end of the interval $(t_0, t_i]$, we define

$$P_i = \sum_{i'=1}^{i} p_{i'} \tag{1}$$

as the total portfolio "recovery value until time t_i", i.e. the total recovery, for all the K debt positions, in the first i periods from the default date.
We define

$$R_i = \frac{P_i}{EAD} \tag{2}$$

as the portfolio "recovery rate until time i", while

$$r_i = \frac{p_i}{EAD} \tag{3}$$

equals the portfolio recovery rate in the i-th time interval of delay $(t_{i-1}, t_i]$.
Since $R_i = \sum_{i'=1}^{i} r_{i'}$ we can refer in an equivalent way to R_i or to r_i, being $r_i = R_i - R_{i-1}$ (for $i = 2, \ldots, I$) and $r_1 = R_1$.

Let's consider the following example.
Consider a portfolio with $K = 4$ debt positions. We are interested in measuring its performances in 3 years after default, i.e. $I = 3$ periods of delay.
The data are in the following table

	I	1	2	3
k	EAD_k	$p_{k,1}$	$p_{k,2}$	$p_{k,3}$
1	100	10	0	0
2	200	20	15	0
3	300	20	25	10
4	400	30	35	10

The portfolio performance can be measured in terms of recovery rates until year i (R_i) as

i	1	2	3
EAD	p_1	p_2	p_3
1000	80	75	20
	r_1	r_2	r_3
	8.00%	7.50%	2.00%
	P_1	P_2	P_3
	80	155	175
	R_1	R_2	R_3
	8.00%	15.50%	17.50%

This means that, for example, in the first 2 years the portfolio recovers the 15.5% of the total initial exposure: 8% in the first year and 7.5% in the second.
Sometimes the available data are incomplete, in particular, censored, because the $p_{k,i}$ is not available from a particular date on. In this case, it is not possible to compute the recovery curve. However, in the next section, we will see how to estimate it from the incomplete data.

3. Estimating the recovery rate curve from censored data

The estimation of the recovery curve in the presence of censored data is carried out in a way similar to the estimation of a survival curve (see for example Kalbfleisch and Prentice, 2002). First, we note that sometimes it is interesting to consider the "conditional recovery rate" c_i in each delay period i. Let E_i be the effective portfolio exposure at the beginning of period i

$$E_i = \begin{cases} EAD & i = 1 \\ \sum_{k=1}^{K} \left(\mathrm{EAD}_k - \sum_{i'=1}^{i-1} p_{k,i'} \right) & i > 1 \end{cases} \tag{4}$$

that means $E_i = EAD - P_{i-1}$ with $P_0 = 0$ by convention.
The conditional recovery rate is defined as

$$c_i = \frac{p_i}{E_i} \tag{5}$$

In words, it is the recovery rate with respect to the effective portfolio exposure at the beginning of the period (E_i) rather than to the initial one (*EAD*).
We observe that it is possible to obtain r_i from c_i and R_{i-1}:

$$r_i = \frac{p_i}{EAD} = \frac{p_i}{EAD} \cdot \frac{E_i}{E_i} = \frac{p_i}{EAD} \cdot \frac{EAD - P_{i-1}}{E_i} = \frac{p_i}{E_i} \cdot \frac{EAD - P_{i-1}}{EAD} = c_i \left(1 - \frac{P_{i-1}}{EAD}\right) =$$

$$= c_i(1 - R_{i-1}) = c_i(1 - \sum_{i'=1}^{i-1} r_{i'})$$

It means that the recovery rate is the conditional recovery of the percentage of how much still has to be recovered.
In our example we have

i	**1**	**2**	**3**
EAD	p_1	p_2	p_3
1000	80	75	20
	P_1	P_2	P_3
	80	155	175
	E_1	E_2	E_3
	1000	920	845
	r_1	r_2	r_3
	8.00%	7.50%	2.00%
	c_1	c_2	c_3
	8.00%	8.15%	2.37%
	R_1	R_2	R_3
	8.00%	15.50%	17.50%

From the previous table we see that the performances of our portfolio are better in the second year rather than in the first one if they are evaluated with respect to the effective exposure.

It is interesting to note that it is possible to compute R_i also in this way

$$R_i = 1 - \prod_{i'=1}^{i} (1 - c_{i'}) \tag{7}$$

because

$$1 - \prod_{i'=1}^{i} (1 - c_{i'}) = 1 - \prod_{i'=1}^{i} \left(1 - \frac{p_{i'}}{E_{i'}}\right) =$$

$$= 1 - \prod_{i'=1}^{i}\left(1 - \frac{p_{i'}}{EAD - P_{i'-1}}\right) =$$
$$= 1 - \prod_{i'=1}^{i}\left(\frac{EAD - P_{i'-1} - p_{i'}}{EAD - P_{i'-1}}\right) =$$
$$= 1 - \prod_{i'=1}^{i}\left(\frac{EAD - (P_{i'-1} + p_{i'})}{EAD - P_{i'-1}}\right) =$$
$$= 1 - \prod_{i'=1}^{i}\left(\frac{EAD - P_{i'}}{EAD - P_{i'-1}}\right) =$$
$$= 1 - \frac{EAD - P_1}{EAD - P_0} \cdot \frac{EAD - P_2}{EAD - P_1} \cdot \ldots \cdot \frac{EAD - P_i}{EAD - P_{i-1}} =$$
$$= 1 - \frac{EAD - P_i}{EAD - P_0} = 1 - \frac{EAD - P_i}{EAD} =$$
$$= \frac{EAD - EAD + P_i}{EAD} =$$
$$= \frac{P_i}{EAD} = R_i$$

being $P_0 = 0$. In the example,

$$R_1 = 1 - \left(1 - \frac{80}{1000}\right) = 8.00\%$$

$$R_2 = 1 - \left(1 - \frac{80}{1000}\right)\left(1 - \frac{75}{920}\right) = 15.50\%$$

$$R_3 = 1 - \left(1 - \frac{80}{1000}\right)\left(1 - \frac{75}{920}\right)\left(1 - \frac{20}{845}\right) = 17.50\%$$

This way of computing R_i is convenient when there are censored data in the database, i.e. for some NPLs the recovery $p_{k,i}$s are observed only until a particular time. In this case, the idea is to apply formula (7) by computing the conditional recovery rate c_i using only the available data. In details, let us suppose that

$$K_i = \{k = 1, \ldots, K \mid \exists\ p_{k,i}\} \tag{8}$$

is the subset of indexes k corresponding to the NPLs for which at delay time i the value $p_{k,i}$ is not censored. In this case the effective portfolio exposure, for $i > 1$, is a generalization of (4):

$$E_i = \sum_{k \in K_i}\left(\mathrm{EAD}_k - \sum_{i'=1}^{i-1} p_{k,i'}\right), \tag{9}$$

and the conditional recovery rate is

$$c_i = p_i / E_i = \left(\sum_{k \in K_i} p_{k,i}\right) / E_i. \tag{10}$$

The recovery rate in the i-th time interval of delay is computed as $r_i = R_i - R_{i-1}$ (for $i = 2, \dots, I$) with $r_1 = R_1$, since formula (3) cannot be used.
Let's consider the previous example where another NPL has been added to the portfolio.

	i	**1**	**2**	**3**	**4**
k	***EAD_k***	$p_{k,1}$	$p_{k,2}$	$p_{k,3}$	$p_{k,4}$
1	100	10	0	0	0
2	200	20	15	0	0
3	300	20	25	10	15
4	400	30	35	10	#N/D

If we want to consider more than 3 intervals of delay, assuming we are interested in measuring the performances in 4 years, i.e. $I = 4$ periods of delay, then

i	**1**	**2**	**3**	**4**
EAD	p_1	p_2	p_3	p_4
1000	80	75	20	15
	P_1	P_2	P_3	P_4
	80	155	175	190
	E_1	E_2	E_3	E_4
	1000	920	845	500
	r_1	r_2	r_3	r_4
	8.00%	7.50%	2.00%	2.48%
	c_1	c_2	c_3	c_4
	8.00%	8.15%	2.37%	3.00%
	R_1	R_2	R_3	R_4
	8.00%	15.50%	17.50%	19.98%

In the example,

$$K_1 = \{ k = 1, 2, 3, 4\}$$

$$K_2 = \{ k = 1, 2, 3, 4\}$$

$$K_3 = \{ k = 1, 2, 3, 4\}$$

$$K_4 = \{ k = 1, 2, 3 \}$$

so that

$$E_1 = (100 + 200 + 300 + 400) = 1000 = \text{EAD}$$

$$E_2 = (100 + 200 + 300 + 400) - (10 + 20 + 20 + 30) = 920$$

$$E_3 = (100 + 200 + 300 + 400) - (10 + 20 + 20 + 30 + 15 + 25 + 35) = 845$$

$$E_4 = (100 + 200 + 300) - (10 + 20 + 20 + 15 + 25 + 10) = 500$$

and

$$R_1 = 1 - \left(1 - \frac{80}{1000}\right) = 8.00\%$$

$$R_2 = 1 - \left(1 - \frac{80}{1000}\right)\left(1 - \frac{75}{920}\right) = 15.50\%$$

$$R_3 = 1 - \left(1 - \frac{80}{1000}\right)\left(1 - \frac{75}{920}\right)\left(1 - \frac{20}{845}\right) = 17.50\%$$

$$R_4 = 1 - \left(1 - \frac{80}{1000}\right)\left(1 - \frac{75}{920}\right)\left(1 - \frac{20}{845}\right)\left(1 - \frac{15}{500}\right) = 19.98\%$$

This method of measuring performances allows not only to measure jointly the recovery rate and the time to liquidate, but also to deal with censored data.

The results would have been different, if we simply did not consider in the portfolio the NPLs for which the data are censored.

In the previous example, with $I = 3$ periods of delay we would had the same results as before, whereas considering $I = 4$ periods of delay we would have not considered NPL_4, obtaining very different results for all the durations considered, as it is shown in the table below.

	i	**1**	**2**	**3**	**4**
k	EAD_k	$p_{k,1}$	$p_{k,2}$	$p_{k,3}$	$p_{k,4}$
1	100	10	0	0	0
2	200	20	15	0	0
3	300	20	25	10	15
4	~~400~~	~~30~~	~~35~~	~~10~~	~~#N/D~~
	EAD	p_1	p_2	p_3	p_4
	600	50	40	10	15
		P_1	P_2	P_3	P_4
		50	90	100	115
		E_1	E_2	E_3	E_4
		600	550	510	500
		r_1	r_2	r_3	r_4
		8.33%	6.67%	1.67%	2.50%
		c_1	c_2	c_3	c_4
		8.33%	7.27%	1.96%	3.00%
		R_1	R_2	R_3	R_4
		8.33%	15.00%	16.67%	19.17%

Obviously, it is wrong to imagine the censored data equal to 0, meaning no inflows instead than no information about that inflow.
With the same example, substituting $p_{4,4} = 0$, we would obtain

	i	1	2	3	4
k	EAD_k	$p_{k,1}$	$p_{k,2}$	$p_{k,3}$	$p_{k,4}$
1	100	10	0	0	0
2	200	20	15	0	0
3	300	20	25	10	15
4	400	30	35	10	0
	EAD	p_1	p_2	p_3	p_4
	1000	80	75	20	15
		P_1	P_2	P_3	P_4
		80	155	175	190
		E_1	E_2	E_3	E_4
		1000	920	845	825
		r_1	r_2	r_3	r_4
		8.00%	7.50%	2.00%	1.50%
		c_1	c_2	c_3	c_4
		8.00%	8.15%	2.37%	1.82%
		R_1	R_2	R_3	R_4
		8.00%	15.50%	17.50%	19.00%

4. Application

We analyze a data set of Italian NPLs supplied by a specialized operator.
We examine one portfolio of unsecured loans with initial debt size between 5000 and 25000 euro (5000 < EAD_k < 25000) and one portfolio of unsecured loans with initial debt size between 100000 and 250000 euro (100000 < EAD_k < 250000). The portfolios have the same year of acceptance by the operator: year 2005. In particular:
Portfolio 1: unsecured, 5000 < EAD_k < 25000, year 2005, $K = 4732$;
Portfolio 2: unsecured, 100000 < EAD_k < 250000, year 2005, $K = 876$.
We consider as time t_0 the year of acceptance (2005), rather than the exact time of default, because that is the moment in which the operator starts the recovery procedure. We follow the recovery history for 9 years.
We observe that both portfolios have approximately 5% censored data in the last year and about 2.5‰ censored data in the penultimate year.
The results in terms of recovery rate (r_i), conditional recovery rate (c_i) and recovery rate until time i (R_i) are reported in the tables below

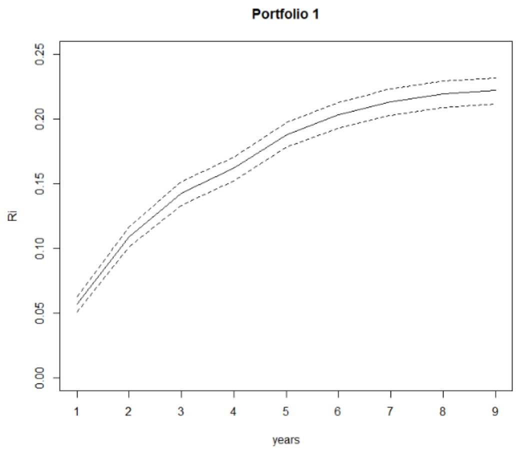
Portfolio 1, EAD=69603519, Average EAD_k=14709

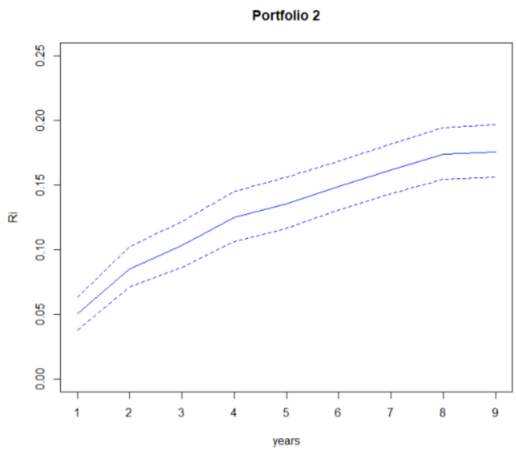
r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
5.69%	5.21%	3.36%	1.97%	2.56%	1.53%	1.02%	0.63%	0.24%
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
5.69%	5.52%	3.77%	2.30%	3.06%	1.89%	1.28%	0.80%	0.30%
R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9
5.69%	10.90%	14.26%	16.23%	18.79%	20.32%	21.34%	21.97%	22.21%

Portfolio 2, EAD=132378119, Average EAD_k=151117

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
5.05%	3.42%	1.83%	2.19%	1.09%	1.29%	1.31%	1.20%	0.19%
c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
5.05%	3.60%	2.00%	2.44%	1.24%	1.50%	1.53%	1.43%	0.24%
R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9
5.05%	8.47%	10.30%	12.49%	13.58%	14.87%	16.18%	17.38%	17.57%

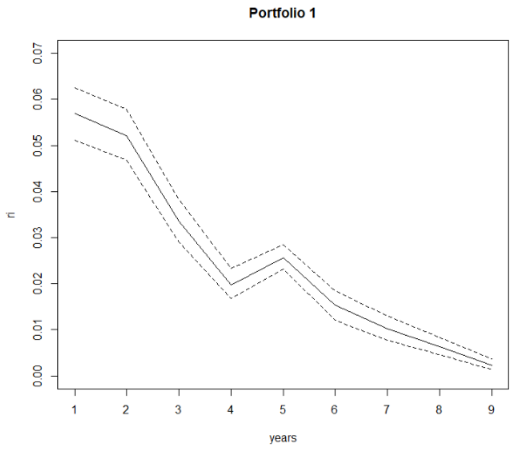
The plot of the results in terms of recovery rate until time i (R_i) is:

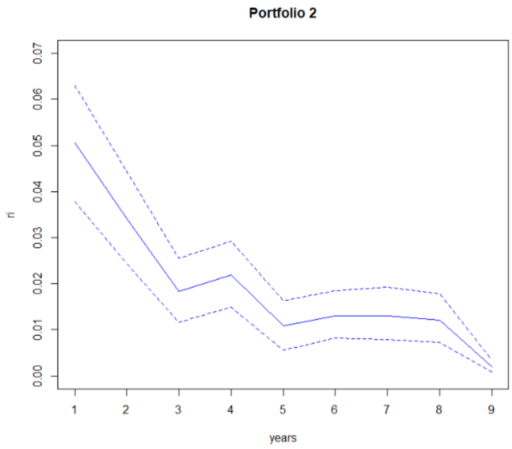




where the dotted lines are the boundaries of the confidence intervals computed pointwise by using a non parametric bootstrap (Efron, 1981).

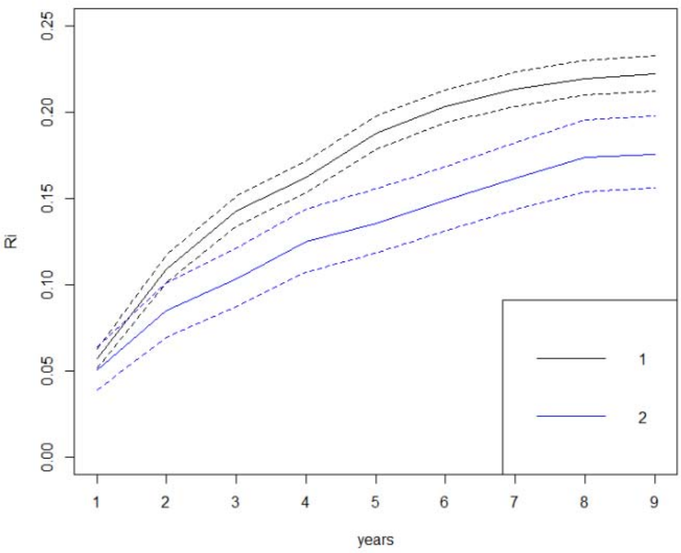
The plot of the results in terms of recovery rate (r_i) is:





Obviously, the highest values of the recovery rate are at the beginning of the period (i = 1) and as time passes the recovery rate tends to decrease, even if not monotonically.

To compare the result is useful to plot together the result of both portfolios. We discuss R_i, that in our opinion is the most explicative ratio.



Even considering the width of the confidence intervals, it appears that the recovery is greater for the portfolio with smaller credits. Probably, this is due to the fact that taking charge by specialized operators has greater effect on those who have to return lower amounts.

5. Conclusions and final remarks

According to the objective of this paper, we propose a kind of measurement that takes in consideration both the recovery rate and the time to liquidate. In our opinion, an efficient way to do that is to measure a "recovery curve" in terms of recovery rate until time i, so to observe the behaviour of the recovery rate during the time.

In doing that, we faced the problem of censored data and we suggest to use a method of measuring performances that allows not only to measure jointly the recovery rate and the time to liquidate, but also to deal with censored data.

This method is based on an algorithm that is usually used in the construction of survival curves.

Our next goal is to extend and test the validity of the method to cases where the database has missing data not only at the end of the observation period, but also at the beginning of it.

References

[1] BIS (2001). *The Internal Ratings-based Approach*. Basel: Bank for International Settlements. https://www.bis.org/publ/bcbsca05.pdf

[2] BIS (2008). *Principles for sound liquidity risk management and supervision*. Basel: Bank for international settlements. http://www.bis.org/publ/bcbs144.pdf

[3] Bruche M., Gonzales-Aguado C. (2010). Recovery rates, default probabilities, and the credit cycle. *Journal of banking and finance*, 34(4), 754–764. https://doi.org/10.1016/j.jbankfin.2009.04.009

[4] Campbell J.Y., Lo A.W., MacKinlay C.A. (1997). *The econometrics of financial markets*. Princeton: Princeton University Press.

[5] Cheng D., Cirillo P. (2018). A reinforced urn process modeling of recovery rates and recovery times, *Journal of Banking and Finance*, 96, 1–17. https://doi.org/10.1016/j.jbankfin.2018.08.014

[6] Devjak, S. (2018). Modeling of Cash Flows from Nonperforming Loans in a Commercial Bank. *Naše Gospodarstvo/Our Economy,* 64(4), 3-9. https://doi.org/10.2478/ngoe-2018-0018

[7] Efron B. (1981). Censored Data and the Bootstrap. *Journal of the American Statistical Association*, Jun., Vol. 76, N. 374, 312-319. https://www.jstor.org/stable/2287832

[8] Ernst&Young (2011). *European non-performing loan report 2011. Restructuring follows strategy – a review of the European loan portfolio market.* Eschborn: Ernst&Young.

[9] Hastie T., Tibshirani R., Friedman J. (2009). *The elements of statistical learning*. Springer.

[10] Kalbfleisch J.D., Prentice R.L. (2002). *The statistical analysis failure time data.* John Wiley & Sons, Inc.

[11] Moody's (2013). *Annual Default Study: Corporate Default and Recovery Rates, 1920-2012.* Retrieved on 5th January 2014 from: http://www.moodys.com/research/Annual-Default-Study-Corporate-Default-and-Recovery-Rates-1920-2012--PBC_151031

[12] Resti A. ,Sironi A. (2007). *Risk Management and Shareholders' Value in Banking*. John Wiley & Sons.

[13] Ye H., Bellotti A. (2019). Modelling Recovery Rates for Non-performing Loans. https://ssrn.com/abstract=3256252 or http://dx.doi.org/10.2139/ssrn.3256252