Frankfurt School of Finance & Management

German Excellence. Global Relevance.

Certified Expert in Risk Management

Unit 4.1: Credit Risk Management

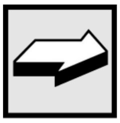
Frankfurt School of Finance & Management

German Excellence. Global Relevance.

Certified Expert in Risk Management

Unit 4.1: Credit Risk Management

Symbols



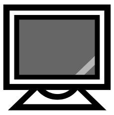
Introduction

Definition

Example

Remember

Further Reading



Video Lecture

6. edition 9/2016

Printed in Germany

Content

Abbreviations

ABC	Activity Based Costing
ADB	Asian Development Bank
ALCO	Asset and Liability Management Committee
ALM	Asset-Liability Management
AML/CFT	Anti Money Laundering / Countering the Financing of Terrorism
ATM	Automated Teller Machine
BCBS	Basel Committee on Banking Supervision
BSC	Balanced Scorecard
CAR	Capital Adequacy Ratio
CGAP	Microfinance Secretariat at the World Bank
COSO	Committee of Sponsoring Organizations of the Treadway Commission
EAD	Exposure at Default
EL	Expected Loss
ERM	Enterprise Risk Management
EUR	Euro currency
FRM	Financial Risk Manager - professional designation
HR	Human Resources
ISO	International Standards Organization
KPI	Key Performance Indicator
KRI	Key Risk Indicator
LGD	Loss given Default
MFI	Microfinance Institution
MIS	Management Information System
MIV	Microfinance Investment Vehicle
MSME	Micro-, Small and Medium Enterprise
NBFI	Non-Bank Financial Institution
NGO	Non-Governmental Organization
NPL	Non-Performing Loan
PAR	Portfolio-at-risk
PD	Probability of Default
PMI	Project Management Institute
PMI-RMP	PMI Risk Management Professional designation
PRM	Professional Risk Manager professional designation
ROA	Return On Assets
ROE	Return On Equity
SME	Small and Medium Enterprise
TA	Technical Assistance
USD	US Dollar
VPN	Virtual Private Network

Learning Outcomes

This Unit 4.1 on credit risk is the single largest unit in the Certified Expert in Risk Management course. It goes to the heart of the matter of risk in financial institutions. The learning objectives are ambitious: We want to teach you everything we know about identifying, measuring, reporting, mitigating and managing credit risk with a special focus on microcredit and SME lending. Once you have worked through the script and all of the exercises, you should be able to:

- communicate effectively about the parameters that determine the loss distribution in credit portfolios.
- understand the nature and drivers of default in MSME credit.
- assemble and interpret descriptive portfolio performance statistics such as arrears schedules, vintage curves or a transition matrix.
- derive appropriate loan loss provisioning methods in compliance with prudential norms and IFRS.
- design collateral strategies for various markets and loan products that are both socially responsible and conducive to minimizing loss given default.
- organize an efficient arrears management and collections process in compliance with responsible finance practices.
- build and maintain statistical credit scoring and rating models for microenterprise credit and SMEs.
- incorporate forward visibility of default probabilities into a risk-based credit pricing framework.

1 Introduction and Overview

Now, we are done with the terminological warm-ups and ready to dig into the details, risk-by-risk. This Unit 4.1 and each of the following Units will take one of the core risk dimensions and walk through the full risk management process from identification to measurement to action. Credit risk is up first. Then comes operational risk, interest rate risk, exchange rate risk, and finally we will deal with the main downstream risk category of liquidity.

We are getting more specific now

Unsurprisingly, this credit risk Unit will be the single biggest piece in the course with the greatest number of analytical tools, the most homework and some of the most interesting number crunching.

We will talk about credit transaction versus credit portfolio risk and about organizational principles of credit risk management in SME lending, micro-enterprise finance and consumer credit. Portfolio risk management always starts with a keen eye on concentrations and the need for effective ex-ante diversification and macro-budgeting of exposures by industry and geography. We will also study traditional portfolio performance diagnostics: such as arrears aging schedules, vintage curves and the transition matrix.

Much time will be spent on the data requirements for predictive credit modeling and the development of a comprehensive client data strategy. This data platform will at the same time also enable targeted marketing and credible reporting on the social and economic development impact of financial access.

Some number-crunching coming up

Assuming that we have available good socio-demographic, financial and credit history data on our clients, we can build statistical models for the probability-of-default and loss-given-default parameters. In fact, you can do this at home with some inexpensive plug-in software for Excel. We will show you step by step how it is done on some examples of real loan portfolio data. The same analytical apparatus will then be used to develop a behavioral scoring model, such as a collections scoring, for example. A collections score will help you decide which clients in arrears might be most responsive to which type of arrears management actions. We will also discuss a detailed borrower and facility rating model for SMEs in emerging and developing markets. And finally, with good estimates for the basic portfolio risk parameters in place, we can now put it all together in a risk-based pricing model.

So let's get started. We repeat our simple definition of credit risk that we introduced in Unit 3:

Credit risk is defined as the possibility that a borrower or other contractual counterparty might default, i.e. might fail to honor their contractual obligations.

Risk of counterparty defaults deferred to liquidity management

Let's keep the **counterparty credit risk** element from the above definition on our check list for later, when we will discuss liquidity risk in Unit 4.5. The most likely context in which an MSME bank would encounter the risk of a wholesale finance counterparty defaulting is in treasury management: investing the liquid asset reserve in bank placements and high grade debt securities issued by government, banks and prime corporates.

Counterparty credit risk aside, we should still amend the above definition by a further dimension called **migration risk**. Of course, credit risk is essentially about the loss that occurs, if a borrower does not pay. Yet, credit losses may arise well before a borrower actually misses a payment. Losses can be triggered simply by the fact that the perceived likelihood of a future default has increased while an exposure is outstanding.

The potential deterioration of the credit quality of an un-defaulted exposure is called **migration risk**. This form of potential loss is generally also subsumed under a broader definition of **credit risk**.

Migration risk is not as abstract as it sounds and even has relevance for MSME credit. Just imagine you have a portfolio of microenterprise loans outstanding in an area where the underlying source of economic activity is a big mining site. Tomorrow, there is an accident at the mine. The main shaft floods, production will be idled for at least six months and most laborers at the mine will be laid off. Nothing has changed in your microcredit portfolio for this month's collections, defaults are low. But clearly you can see the train wreck coming. If you happen to routinely put microcredit receivables into a security pool for multiple lenders or even sell loans into a securitization vehicle, these future credit losses will be immediately monetized. A reasonable counterparty would value this loan portfolio much lower on the day after the mining accident. Thus a bigger valuation haircut will be imposed as you borrow against the loans, and selling them might become entirely impossible.

Migration risk is important for pricing longer-term loans

We will come back to the idea of **migration risk** as we consider the pricing of longer-term loans in Chapter 10 of this Unit. If we make loans for several years, we only get one chance up front to assess the risk of default and price it into the loan. Much may happen that would increase the risk of default over time. Thus, we should anticipate a certain downward migration of the credit quality and factor in an additional charge for credit migration risk.

Naturally, from the perspective of a microfinance institution or a traditional retail / SME bank, the big deal when it comes to credit risk is the short-term **borrower default component.** The "simple" question of whether or not the borrower will default over the next year and how much we stand to lose, if he does, is what we will look at under the microscope from all different angles now.

As we zoom in on the risk of small businesses and individuals failing to meet their loan obligations, we will continue to use the distinction between the transaction risk and portfolio credit risk dimensions:

Transaction risk refers to individual loans and essentially measures (1) the standalone probability that the borrower will be able to repay, as well as (2) the ultimate loss in the case of a borrower default after use of collateral and other mitigating factors.

Portfolio credit risk is concerned with measuring correlations between individual borrower defaults, the effects of diversification, the cyclicality of collateral values and the implications of reputation and contagion effects in microcredit.

There is so much to say, so many stories to share, so many models to discuss, when it comes to the default risk of micro- and SME borrowers. Where do we start and how do we bring some conceptual order to all of this?

Use expected loss formula as conceptual framework for credit risk

We propose to use one simple formula as our guiding light through the entire Unit. You have probably seen this before - it has been made famous by the Basel II rules on regulatory capital for credit risk.[1] The financial crisis has dented the confidence in credit risk modeling a bit. But nonetheless, this basic little formula was not invented by the Basel Committee and it is also not discredited by some of the shenanigans that banks have used to keep their capital requirements low. It is as simple and always true as saying the 'world is round':

Expected Loss = Probability of Default * Exposure at Default * Loss Given Default

Or, for short we will write: EL=PD*EAD*LGD.

By itself this "model" does not explain anything, of course. It is just a very simple way to structure our thinking about the losses from borrower default into three elementary dimensions that we can take apart in more detail and develop specific models for. So, instead of just saying "credit risk is when the guy does not pay the loan back", we look at this loss as arising from three distinct factors:

1) There must be a **default**. This is an event or behavior which is a characteristic of the borrower.
2) The loss then depends on how much the client owed the particular lender when he stopped paying, i.e. the **exposure at default**.
3) And finally, we need to factor in **how much of that exposure will actually be lost** after we liquidate any collateral and attempt to collect through the legal and arrears management routes.

[1] See paragraphs 211ff in BCBS 2004 Basel II (June 2006 compilation).

2 The EL=PD*EAD*LGD Way of Thinking

Let us unpack the EL=PD*EAD*LGD logic further. For this, we first need some more specific definitions:

We define the **Expected Loss (EL)** as the average or mean amount of credit loss to be incurred over a particular time period. The loss is measured as the present value or book value of receivables that will not be collected or will have become unrecoverable and therefore will be written off or otherwise expensed during a particular period of time.

The **Probability of Default (PD)** is the percentage probability of a borrower entity to produce a default event as perceived by the lender over a specified period of time, typically one year. The PD is most often stated for a future period beginning immediately, but can also be expressed as a forward default probability beginning in one year for one year, for example.

The **Exposure at Default (EAD)** is the total balance owed by the borrower to the particular lender at time of default expressed in currency units.

Loss-given-default (LGD) is the percentage of EAD that is considered lost, once it has been established that a default has occurred. The LGD is equal to 100% minus the percentage of EAD that will be recovered by way of liquidation of collateral and other post-default collection actions. For the purposes of establishing LGD, the post-default cash flows from recoveries must be discounted back to the time of default at the original internal rate of return of the defaulted contract.

Figure 1 illustrates the relationship between the LGD and the net present value of post-default recoveries. IAS 39 requires that receivables for which evidence of impairment exists be carried on the balance sheet at the net present value of residual realizable cash flows. The discounting should be done at the effective rate (internal rate of return) of the original loan contract. Obviously, materialized default is a very clear "evidence of impairment", so the impaired receivables valuation under IAS 39 applies. We will get back to LGD, collateral, IAS 39 / IFRS 9 and impairments and provisioning in more detail in chapters 5 and 6. We just wanted to give an initial understanding of LGD here, so we can appreciate the EL=PD*EAD*LGD logic.

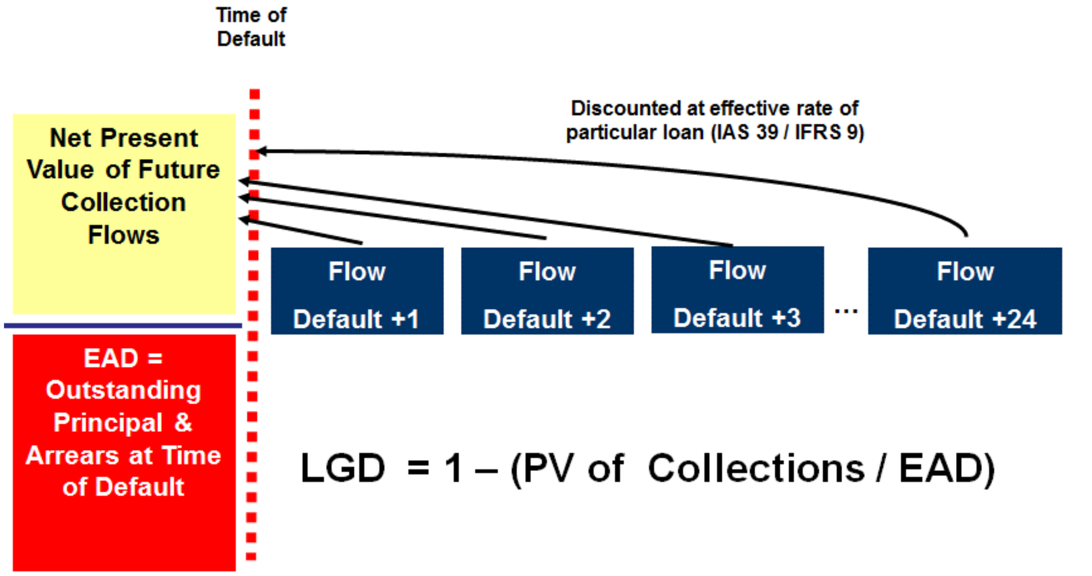


Figure 1: Definition of LGD and Relationship to Present Value of Recovery

How to estimate EAD

We should also expand a bit on the need for a separate **EAD dimension**. After all, is the amount disbursed or the amount outstanding today not the obvious measure of exposure? Absolutely, the amount outstanding today is probably a good starting point for how much gross exposure is at risk, should the particular borrower default. However, the point is that we want to establish an expected loss for a forward period of time, say for a year starting tomorrow. Since the default may occur at any time over the next year, it is obvious that the amount outstanding at the time of default is also an uncertain "random" variable that has some relationship with the amount outstanding today, but by no means must be equal to it.

Example: EAD. Let's imagine a simple installment loan that is reimbursed in 12 equal monthly installments annuity style. The loan is disbursed today and we wonder how much might be outstanding, if and when the borrower defaults over the course of the coming year. If we know or assume that the borrower has a PD of 3% for the year, we could surmise that default is equally likely throughout the year and that the borrower would default with 3%/12 every month and that on average, the default would occur just after the sixth installment due date. If it was a loan with linear principal reimbursement, the expected EAD at that time would be 50% of the starting balance. On annuities, the principal would still be higher, of course. Exactly 52.5%, if the loan had a nominal annual rate of 20%.

Exercise M4.1_Ex1: Annuity loan formulas

We can't let this opportunity pass to do another little finger exercise in annuity formulas in Excel: What is the principal balance outstanding at the end of month 5 (just after receipt of the 5th installment) for an annuity style loan of 1,000 that is reimbursable in 18 equal monthly installments and that carries an annual nominal rate of 24% (i.e. 2% per month)?

Hint: remember the annuity formulas =PV(), =PMT(), =PPMT(). Also consider that the principal balance is always equal to the present value, =PV(), of the future loan payments discounted at the loan rate.

Solution: 756.96. See M4.1_Ex1_Annuity.xlsx for details.

EAD < current exposure is not plausible

Back to the idea that the default might occur on average half way through the year with an exposure of default that has been diminished by six interim principal installments. Sorry, to have led you on. We did learn a new trick in Excel along the way maybe, but this idea of EAD being significantly smaller than the current exposure outstanding just does not hold water. Not even in the simplest microcredit cases, where we only make fixed installment loans for tenors of around one year, one client, one loan at a time. In fact, we venture say that as a reasonable assumption we should go with the notion that **EAD will always be larger than the current un-defaulted balance outstanding** for the particular borrower. Un-defaulted means that there are no pre-existing arrears on this loan. Here is why EAD tends to be larger, not smaller than the balance outstanding today:

Pre-default liquidity stress tends to maximize EAD

1) You could say, if we are only considering currently un-defaulted exposures and default might be commonly defined as reaching 90+ days in arrears, well than **nobody can default in the first three months of the observation period, because they could not have reached 90+ days in arrears.** And thus the balance outstanding would have been reduced by three principal installments in the meantime. But this argument is not true, because the client could certainly stop paying with the first installment due in the observation period. By the time he reaches default, he would then precisely not have made those three installments. And that means that the balance outstanding when we finally register the default at 90 days has not diminished at all. Instead it increased by the interest not paid with the previous missed installments.

2) When corporations go bankrupt, they always go down in flames with the last dollar on the last credit line drawn to the limit. Think of Enron or Worldcom and other famous disasters. **Default is by definition characterized by a shortage of liquidity and the desperate search for fresh credit**, until the lenders pull the plug and say: "Not a cent more".

3) The corporate example shows nicely that the **balance outstanding profile over the course of the next year, is not independent of the default variable**. If the client takes the default bullet from the Russian roulette, so to speak, we already know that the balance outstanding profile will have evolved differently until the time of default, than if all had been going well for the company. **Default maximizes the balance outstanding profile.** Thus, EAD can always be assumed greater than current exposure. How much greater, naturally depends on institutional practice and lending policies, the prevalence of open limits and overdraft lines etc. It must be studied and modeled empirically.

EAD > current exposure also applies in microcredit

4) The corporate credit logic of **EAD being greater than current exposure actually transfers seamlessly to microcredit**. Also in microcredit, a default will be preceded by a deterioration of the borrower's financial situation. There will be attempts to make up for the liquidity shortage by borrowing more: from your institution, from a competitor, from family or from the village loan shark. Now, think of the prevalent **practice of early settlement and re-advance in micro-lending**. For the most part, these are good borrowers requiring more cash more quickly to grow their business. But if you are not watching it closely, the future defaulters who are already short on cash because their business is failing, might be using the same process to top off their loan balance with the MFI one last time, before they will never pay again. And even though in microcredit, it is not common practice to grant multiple parallel loans or revolving lines of credit, many institutions do offer emergency and other short-term special purpose loans in parallel to the main working capital product. Et voilà, what could be a more obvious emergency than failing in the business and needing cash urgently.

5) Finally let's consider the **effect of the "graduation principle"** in microcredit. This means that borrower relationships are built up gradually over time with ever increasing loan amounts in each loan cycle. So, even if all goes as planned and there is no deterioration of the financial situation of the client, it is quite likely, that over the course of a year, the current loan would have been entirely paid off and replaced by a new larger loan. The new loan would be early in its maturity and most likely display a higher balance outstanding than we see on the current loan at this time.

Thus all told, even in microcredit, **EAD will typically exceed current exposure** to the particular lender.

Now, that we have a bit of a feeling for the nature of the three components in the EL formula, let's look again at the whole picture of EL=PD*EAD*LGD.

EL=PD*EAD*LGD as Random Variables

On the day that we disburse a loan and look ahead at the next year, all three factors (PD, EAD, LGD) are what we call random variables.

In probability and statistics, a **random variable or stochastic variable** is a variable whose value is subject to variations due to chance. As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown). Rather, it can take on a set of possible different values, each with an associated probability.

A random **variable's possible values** might represent the outcomes of a yet-to-be-performed experiment or an event that has not happened yet, or the potential values of a past event whose already-existing value is not yet known.

A random variable can be classified as either **discrete,** i.e. it may assume any of a specified list of exact values, or as **continuous**, i.e. it may assume any numerical value in an interval or collection of intervals.

A discrete random variable that can take one of a limited, and usually fixed, number of possible values is called a **categorical variable.** A categorical variable that can assume exactly one of two possible values (e.g. [yes; no] or [0; 1]) is termed separately as a **binary variable** or **dummy variable**.

The mathematical function describing the possible values of a random variable and their associated probabilities is known as a **probability distribution.**

PD, EAD LGD at loan versus portfolio level

The elements PD, EAD and LGD each are random variables for every individual loan. EL as a function of these three random variables then **is also a (derived) random variable.** Its value depends on the outcome of each of the three underlying random variables.

At the same time, we can think of EL and its components as random variables at the aggregate portfolio level.

Just as EL_i (i.e. the EL for an individual loan) is a random variable that is derived from the outcomes of PD_i*EAD_i*LGD_i, we can look at EL_p as an aggregate random variable EL_p=PD_p*EAD_p*LGD_p that is the result of the summation of individual loan losses to the portfolio level.

Let's experiment with these random variable concepts a little in Excel. The notion of default is a binary random variable that is most frequently coded like this: no default = 0, default =1. The probabilities of the two outcomes are expressed typically for a specific period of time, often for a year. The specific time horizon is obviously necessary, otherwise the question would resemble the long-run probability of death, which is 100% always.

Imagine a borrower who has a PD of 5% per annum. This means he would default 5 times in 100 years. Or, among 100 identical and independent borrowers each with a 5% PD, in an average year, five would have defaulted and 95 would still be in good credit standing at the end of the year. Yet, in loan default, there is not just one "lottery drawing" at the end of the year. Instead, borrowers play a version of Russian roulette, whereby the revolver is spun at least 12 times a year, or every time that a loan installment is due. We don't want to overstretch the Russian roulette analogy, but it really explains it quite well: with multiple elimination rounds in a year, what will be the odds of taking a bullet at each round, so that at year-end you would end up with 5 defaulters per 100 borrowers? In the default/no-default game, just like in Russian roulette, only survivors get to spin the revolver again. So, we can say that the 5% 1-year PD equates to a 95% cumulative survival rate after 12 monthly elimination rounds. With this, we can find the equivalent default rate at each monthly elimination such that:

Period equivalent default rates

$$(1\text{-}PD_{monthly})*(1\text{-}PD_{monthly})\ldots *(1\text{-}PD_{monthly}) =(1\text{-}PD_{annual})$$

$$PD_{monthly}=1-(1-PD_{annual})^{(1/12)}$$

This is true because conditional probabilities are concatenated by multiplication: Only on the condition that you did not default in the first monthly round, do you get to spin the revolver again. And only the survivors of the second round get to play the third game etc.

If you are not familiar with the exponent notation in Excel, the ^ sign means "to the power of". And to the power of 1/12 is equivalent to taking the 12th root of something. So, have you typed it into Excel, yet? The monthly default probability that would be equivalent to a 5% annual default rate is: 0.427%.

Please keep Excel open, we want to look at the expected loss of a single loan under the relationship EL=PD*EAD*LGD. We will use the random number generator in Excel to do a simulation of the loan loss.

Exercise M4.1Ex2

We copied in above the first line from the worksheet M4.1_Ex2_DefaultRate. The Loss is the result of multiplying the binary default variable with the EAD and LGD values. All three are set up as random variables.

Borrower No.	Loan Balance Outstanding	DefaultYes=1	EAD	LGD	Loss
1	1,000.00	1	1,139.52	0.485773	553.55

Figure 2: Screenshot of Datasheet in M4.1_Ex2_DefaultRate

Here is how we did this: The function =RAND() produces realizations of a random variable that are distributed with equal probability in the interval [0;1]. So, in order to get 5% defaults and 95% no defaults, we should write =IF(RAND()>0.05,0,1). All =Rand() values are refreshed with a new random result every time you save or recalculate the spreadsheet. You can trigger a refresh manually with the F9 button.

So, please play with it a little and watch how the results change as you keep pushing the F9 button.

We also want to set up **EAD as a random variable**. Suppose we know that EAD should vary in the range of 110% to 130% of the amount currently outstanding. If the current balance is in cell B2, we can write =(1.1+RAND()*0.2)*B2.

For the **LGD random variable**, we got a bit more fancy: assume that we know from many years of lending experience to this type of market and with this type of collateral that LGD varies narrowly around the 60% average. This would be a case for using a normally distributed LGD variable with a mean of 60% and a narrow standard deviation of just 10% up and down from the 60% mean. We can generate random realizations of such a LGD variable by putting the equally distributed random numbers from RAND() through the inverse normal distribution function. This is done such that the RAND() value becomes the probability, for which we look up the limit along the x-axis that will give us that percentage value of cumulative probability in the normal distribution:

LGD =NORMINV(RAND(),0.6,0.1)

Experiment with the formulas and use Excel help, if you need more explanations

However, this formula could occasionally produce LGD values above 100% and below 0%, which make little sense. We should therefore cut off any excess over 100% or below 0% by using the =MIN() and =MAX() functions in Excel. That's why in the cell E2 we wrote:

LGD =MIN(MAX(NORMINV(RAND(),0.6,0.1),0),1)

In cell F2 you then find the **random loss amount** for the single loan as a function of the Default EAD and LGD random variables. If you push F9 100 times you will get a non-zero loss about 5 times.

Now copy and paste line 2 another 999 times below the first loan to get one thousand identical and independent individual loans that all follow the same random behavior of annual loss.

You can count the total number of defaults in the portfolio and obtain the random default rate for the portfolio in H2 as:

Portfolio Default Rate = SUM(C2:C1001)/1000

As you keep refreshing the random variables, you will see that the portfolio default rate varies around the 5% mark. Thus the PDs of the individual loans add up straight through to the portfolio level. The **portfolio probability of default is the arithmetic average of the individual PDs of each loan**.

Try it: Change the PD in the first 300 loans to 3%. We would expect a portfolio default rate of: (300*3%+700*5%)/1000 = 4.4%. And indeed, the Portfolio default rate now varies around the 4.4% mark as we push F9 repeatedly.

Portfolio expected loss is equal to the sum of loan level expected losses

More generally speaking, the Excel simulation shows that the **expected losses from the individual loans simply add up to the expected value of the portfolio loss variable**.

However, for the other "moments" of the probability distribution, particularly the standard deviation of the portfolio loss distribution, we cannot just add the individual values to find the portfolio-level equivalent.

Referring back to our definitions of risk in Unit 1, you will remember that the **average expected portfolio loss is not "risk"**. This level of loss is pretty much a certainty and should be priced in and charged to borrowers by way of a mark-up on the interest rate. **The measure of risk is in the deviation from that expected average loss.** One should be concerned about the risk of having a bad year where portfolio losses are 10 times the expected value. This risk can be measured by the average deviation from the average portfolio loss, which is called the Standard Deviation.

In statistics and probability theory, **standard deviation,** often represented by the symbol sigma σ, shows how much variation or "dispersion" exists from the average (mean, or expected value). The standard deviation of a random variable, statistical population, data set, or probability distribution is the square root of its variance.

The standard deviation of a probability distribution can be estimated by the standard deviation S on a sample of N observations as:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

The formula includes the common **sample bias correction** in that we take the average of the squared deviations from the mean by dividing by (N-1) and not by N. This is because we already calculated the mean on the same sample, hence the degrees of freedom are N-1. The sample standard deviation is available in Excel with the formula =STDEV()

Subadditivity of risk

Coming back to the **standard deviation of aggregate portfolio losses** as the measure of credit risk: Consider our loan portfolio in M4.1_Ex2. Instead of 1,000 loans with $1,000 each outstanding, imagine we just had a single loan to a single borrower with a balance of 1 million and all else equal: same PD, same EAD and LGD random variables. The expected loss is the same: 5%* 1.2 * 1 million * 0.6 = 36,000. Yet, the standard deviation around this expected portfolio loss will be smaller, if the portfolio is broken down into 1,000 individual loans rather than one big one. This is the famous subadditivity of risk and a result of diversification. This simply means that in a portfolio all borrowers will never default (or not default) at the same time, which prevents extreme deviations from the expected incidence of default. **We measure diversification in a portfolio by the degree of correlation**. Correlation is essentially a measure for the effect that if one defaults, how many of the other 999 in the portfolio will also default.

Effect of diversification & correlation

Under **perfect "positive" correlation**, all 1,000 loans will always default or not default together. This is identically risky as having lent the full one million to just one borrower, hence the standard deviation of the portfolio loss would be the sum of the standard deviations of the individual loan losses. In every other case, other than perfect positive correlation, the standard deviation of the portfolio loss will be smaller than the sum of the individual standard deviations. This is how diversification reduces risk.

Now, we would like to demonstrate the effect of portfolio diversification in an Excel simulation. We start with the same elementary loan as before:

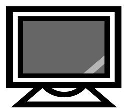
Exercise M4.1_Ex3

- Balance outstanding: 1,000
- PD: 5% p.a.
- EAD: equally distributed in the interval [1.1; 1.3] * balance outstanding
- LGD: Normally distributed with mean 60% and standard deviation 10%, values below 0 and above 100% replaced by those limits

In order to keep it within the line limitation for Excel 2003 users, we copy that elementary loan with the underlying random variables 64,000 times into the same sheet. We will group these 64,000 loans into 128 parallel observations of the same portfolio consisting of 500 loans. Every 500 lines we sum the realized losses of the 500 preceding loans. That gives us 128 random snapshots of the aggregate portfolio loss with the benefit of diversification across 500 loans.

Loan Balance Outstanding	DefaultYes=1	EAD	LGD	Loss	Portfolio Losses	STDEV Portfolio Losses	STDEV Indiviual Loan Loss *500
1,000.00	0	1,253.37	0.688478	-		3,327.82	79,056.57
1,000.00	0	1,155.09	0.490701	-			
1,000.00	0	1,149.76	0.620816	-			
1,000.00	0	1,288.84	0.601142	-			

Figure 3: Screenshot from M4.1_Ex3_PortfolioLoss.xlsx



The standard deviation among these 128 observations of portfolio loss is in the $3,600 range, exactly 3,327.82 in this particular snapshot above. In parallel, we also calculated the standard deviation across all 64,000 individual loans along the column "Loss". The per loan standard deviation varies around the $160 mark. If you added these standard deviations up for all 500 loans in the portfolio, you would get a sum of standard deviations of around 80,000, i.e. 79,056.57 in this particular observation. The sum of standard deviations in I2 is more than 20 times larger than the actual standard deviation of the diversified portfolio loss in cell H2. This is subadditivity at work.

The way we set up the portfolios in PortfolioLossM4.1_Ex3 is actually a very specific situation among the possible diversification scenarios. The defaults on each of the 500 loans in the portfolio are **entirely independent from each other**, such that the outcome (default/no default) on one loan is not at all influenced by the fact that any of the other loans have defaulted or not. Perfect independence like this means entirely uncorrelated outcomes, i.e. a correlation coefficient of zero.

Follow along in Excel: M4.1_Ex4_Portfolio-Correlation.xlsx

In our next exercise **M4.1_Ex4_PortfolioCorrelation,** we use the same set-up of 128 observations on a loan portfolio of 500 loans as in M4.1_Ex3. In order to make the file smaller and concentrate just on the impact of the default variable, we replace the EAD and LGD random variables by their fixed expected values. Now, instead of independence between the loan defaults, we assume perfect positive correlation, i.e. a correlation coefficient of 1.

Perfect positive correlation is easy to simulate. If one of the 500 loans in the portfolio defaults, they all default. So, if we know the outcome of the default variable for the first loan, we know the result for the next 499. Hence, we set the default variables in the 499 other loans equal to the variable in the first loan. This set-up we repeated 128 times, to simulate the outcome of 128 independent years for this perfectly correlated portfolio. And here, no surprise, the risk is fully additive: the standard deviation of the portfolio losses is equal to the sum of the standard deviations of the losses on the 500 individual loans. In the snapshot below, 82,174,55 is roughly equal to 81,853.57. If you keep pushing F9 on the sheet Positive Correlation in M4.1_Ex4, you will see that the two observed standard deviation values in the cells H1 and I1 always come out close to each other.

0	Loan Balance Outstanding	DefaultYes=1	EAD	LGD	Loss	Portfolio Losses	STDEV Portfolio Losses	STDEV Indiviual Loan Loss *500
1	1,000.00	1	1,200.00	0.6	720.00		82,174.55	81,853.57
2	1,000.00	1	1,200.00	0.6	720.00			
3	1,000.00	1	1,200.00	0.6	720.00			
4	1,000.00	1	1,200.00	0.6	720.00			
5	1,000.00	1	1,200.00	0.6	720.00			
6	1,000.00	1	1,200.00	0.6	720.00			
7	1,000.00	1	1,200.00	0.6	720.00			
8	1,000.00	1	1,200.00	0.6	720.00			

Figure 4: Screenshot from M4.1_Ex4_Portfolio Correlation.xlsx

Imagine a portfolio of pairwise negatively correlated borrowers

In the second sheet in the same file for M4.1_Ex4, we now try to simulate a **strongly negatively correlated portfolio** of 500 loans, again with 128 parallel observations of the same portfolio. This is a little tricky to do in Excel, but we think we figured out a reasonable approximation. Each one of the 500 loans in the portfolio still has a standalone PD of 5% p.a., but this time the portfolio displays a high degree of compensating negative correlation. This would be the case, if for each loan that we make, we try to find another borrower who is in an opposing business, which does best when the first business defaults. Imagine always making one loan to a major corporation and another to a bankruptcy law firm headquartered in the same town. Both have a relatively low standalone PD and are definitely rather negatively correlated. If the corporation goes bankrupt, the bankruptcy lawyers are definitely not going to default. If the lawyers default, it is probably because corporate business in the town is doing very well and there have not been any major bankruptcies in the area in a long time.

In exercise M4.1_Ex4, **sheet Negative Correlation**, we tried to implement this idea as follows: We look at the portfolio as 250 pairs of opposing, negatively correlated loans. All have a standalone PD of 5% but loan 1 is matched with loan 251, loan 2 is matched with loan 252 etc. If borrower 251 defaults, borrower 1 certainly has a good year and does not default. This should also work in the other direction, but if you type the reverse relationship into the same formula between the two loans you naturally get a circular reference. So, for the reverse relationship we just shifted the pairs by one loan: the reverse effect is now between loan 1 and loan 252, loan 2 and loan 253 etc.

The default function in Cell C2 for Loan 1 is

=IF(C253=1,0,IF(RAND()>0.05,0,1))

and the reverse reference is found from cell C252 onwards for loan 251 as follows:

=IF(C2=1,0,IF(RAND()>0.05,0,1))

In plain English, the content of cell C252 just above says: if the paired loan number 1 defaults, this loan number 251 will certainly not default, otherwise loan 251 can still default with 5% PD. If we copy this reference pattern among the 250 loan pairs across each of the 128 portfolios in the worksheet, we can see that there is a **further reduction in the standard deviation of the portfolio losses** relative to the portfolio losses in the base case with independent defaults.

In the screenshot below from M4.1_Ex4 the standard deviation of the 128 observed portfolio losses is 2,987.95 versus 3,501.06 in the independence case and 76,418.92 under full positive correlation.

Loan Balance Outstanding	DefaultYes=1	EAD	LGD	Loss	Portfolio Losses	STDEV Portfolio Losses	STDEV Indiviual Loan Loss *500	Number of Defaults in 64,000 Loans
1,000.00	0	1,200.00	0.6	-		2,987.95	76,418.92	4.73%
1,000.00	0	1,200.00	0.6	-		Compared to Independence:		
1,000.00	0	1,200.00	0.6	-		3,501.06		
1,000.00	0	1,200.00	0.6	-				
1,000.00	0	1,200.00	0.6	-		**Perfectly negative Correlation**		
1,000.00	0	1,200.00	0.6	-		Correlation Coefficient:	-100%	
1,000.00	0	1,200.00	0.6	-				

Figure 5: Screenshot from M4.1_Ex4_PortfolioCorrelation, sheet Negative Correlation

Now, that we have a general idea about the "EL=PD*EAD*LGD way of thinking", we should take a more detailed look at the real world of microcredit and SME lending. PDs are not just simply "known" like the odds of rolling a six in a game of dice. Most of this Unit is about how, when and why a borrower default might occur, so that we can better estimate the probability of default at time of making the loan decision. We will also have another chapter about the finer details of LGD estimation and collateral, but understanding default is the big deal in credit risk, of course.

Chapter 2 - Review Questions
1. Define the expected loss in loan portfolio. What are its components? 2. All else equal, will increasing the effective rate charged on a microcredit product tend to increase or reduce LGD? 3. Name two reasons why EAD > balance outstanding is a plausible assumption in microcredit. 4. Default is generally modeled as what kind of a random variable: continuous, binary, or dummy variable? 5. How do you convert an annual PD into an equivalent calendar quarter PD? By dividing by four? 6. What does the subadditivity of risk refer to: the addition of expected losses in a portfolio or the aggregation of the standard deviation of the losses?

3 The Nature of Default

So far we have not really defined default, other than implying very generally that:

Default it is a time-specific characteristic or behavior of a borrower who is no longer meeting contractual obligations under a loan agreement.

Wikipedia offers the following similar definition of default: "A default is the failure to pay back a loan. Default occurs when a debtor has not met his or her legal obligations according to the debt contract, e.g. has not made a scheduled payment, or has violated a loan covenant (condition) of the debt contract."

The Wikipedia definition is largely similar to what we intuitively had in mind, but **it lacks the time aspect**, which we think is important. We should be able to pin down the particular moment when default occurred. The credit risk of a borrower may have been increasing gradually for a while, making a default more likely, and then finally, on a particular day, the default materializes.

In corporate lending, the classic case of a distinct default date is a bankruptcy filing. The nature of a bankruptcy is the formal admission by the entity that it will not be able to meet all of its contractual obligations (insolvency). It therefore relinquishes control of its finances to a court supervised process that tries to allocate residual payments to creditors in proper order and proportion.

Nature of default under liquidity constraints

In consumer credit and MSME lending, and particularly in developing countries, a formal bankruptcy filing is a rare exception. Most often, default manifests itself in missed payments on scheduled loan obligations. We call a delay in loan payments relative to a contractual schedule **an arrear**.

Yet, not every small or temporary arrear immediately constitutes a default. Arrears are a symptom of liquidity constraints at the borrower.

Major corporates in developed markets, even very credit-risky ones that might default in the near future, typically are not facing such binding liquidity constraints: Enron Corporation always paid on time everything that was contractually due, until it filed for bankruptcy on 27 February 2001 and never paid again.

Arrears as a credit risk indicator

In retail and MSME finance, however, liquidity constraints rule. Even a medium-sized business in a developing market may have such limited access to credit that it might miss a loan installment, simply because an incoming cash flow from a major sale is delayed for a few days. Again, such a temporary arrear should not immediately be defined as a default. But its occurrence is nonetheless a warning sign. If repeated arrears arise and their duration grows, then obviously the business is overextended, the average cash reserves have declined and/or the volatility of revenues has increased. This is why arrears are frequently used in retail and MSME credit as a credit risk indicator, i.e. as a warning sign of increasing probability of default.

In lending to corporates that have a professional finance function and lines of credit, arrears would be worthless as an early warning sign. There, a contractual arrear is simply the final manifestation of default.

The notions of liquidity constraints and arrears as credit risk indicators are also **extremely relevant in microenterprise credit** in developing countries. The reality of lending to poor, economically active clients is such that there is never enough disposable cash to meet all loan obligations and all legitimate basic needs of the household. And we mean not just during a crisis, or if the economic condition of the borrower has deteriorated after loan disbursement. By definition, poverty means that not all basic nutritional, health, education and shelter needs of all household members are being met, even when the economic activities of the household develop normally. A child will go hungry, a school fee will not be paid, a sickness will go untreated, a moral obligation to a relative in distress will not be honored, every single week.

Poor households face extreme liquidity constraints daily

This is a critical message for managing credit risk in microfinance: **There is never enough cash to go around and meet every important need of the household.** The only way to get the loan installment paid then is to make sure that the borrower prioritizes paying the installment over other immediate needs. Loan installments only remain high on this priority list, if ...

- the loan is deployed in an income generating activity and the loan size is very carefully calibrated to what can be profitably absorbed by the activity,
- the loan is seen as a ticket to gradual **progress out of poverty**,
- borrowers understand that only by paying, they will **maintain access to credit** at increasingly favorable conditions,
- the lender stays in touch and **monitors the client**,
- **an affinity develops** for the lender as a socially conscious financial service provider, who truly cares about the economic progress of the borrower.

These few bullet points drive much of what we call **best practice in credit risk management for microfinance**.

When do arrears turn into manifest default?

Coming back to the **arrears versus materialized default** discussion: Should the economic situation of a poor borrower actually deteriorate, then liquidity problems will worsen and we should expect to see more frequent and longer lasting arrears. At some point, this behavior is no longer just a warning sign that indicates higher credit risk. When arrears accumulate to several months of unpaid installments, the **arrears become the practical manifestation of default**.

Where to put this arrears threshold that we define as materialized default? Typically, default is declared when a borrower reaches anywhere between 60 to 180 days in arrears. With monthly installments, this would mean having a backlog of three to six unpaid installments. Lenders generally declare the default at a level of arrears, where they are no longer interested in the continued business relationship with the borrower. That is the time when the legal demand letters get sent out and an account may be handed over to a specialized collections team. At default, the dominant motivation is loss minimization. This implies getting as much back from the client as possible by means of drawing on guarantors and deploying various other collection actions.

Parallel alternative definitions of default may exist at the same lender for particular purposes. In credit scoring, for example, the "bad loan" definition is always a key parameter in calibrating the predictive performance of the model. See chapter 7.

Drivers of Default in MSME Credit

The key to managing credit risk in MSME lending is obviously a succinct understanding of the reasons why micro and small business borrowers might stop meeting their loan obligations. Consistently honoring one's contractual obligations requires a **combination of willingness and ability to pay**. Where one or both of these elements are lacking, default occurs. Under the constraints faced by low-income or poor households in developing / emerging markets, willingness and ability to pay are often indistinguishably intertwined.

Ability and willingness to pay are interdependent in microcredit

Ability to pay rarely is as clear-cut as approaching the loan payment date with all other bills paid, all household needs met and a few hundred dollars emergency cash left over after making the loan payment. If that is the required ability-to-pay standard, no microloan and hardly a small business facility will ever be disbursed. Instead, **ability to pay almost always comes down to willingness to pay**, i.e. the discipline to make sacrifices and to prioritize the loan installment over many other competing needs and wants.

Therefore, the single most important factor in assessing the credit risk in consumer loans, as well as in micro-enterprise and small business exposures, is **determining the moral character of the client**. The moral integrity serves as a **proxy for the willingness to pay even in the face of hard personal sacrifices.**

There are **many other proxies for the client's earnest willingness to pay** that we routinely try to assess in the credit decision process. We look for factors that indicate stable life circumstances and responsible behavior: being married and caring for children, or simply being a woman rather than an impulsive and irresponsible male, are all good for stability.

Proxy indicators for the willingness to pay

Having some assets, owning land or a home, or having a fixed-line telephone as opposed to just mobile phones can also be predictive of responsible, disciplined behavior. Again, when in microcredit in developing countries we ask the question "Do you own or rent the home you live in?", we are less interested in the collateral value that could be realized by evicting the family from the home and selling it. Very often, that is not legally feasible and there is no market for self-built structures on communally owned land. However, if a poor household at least owns the modest shelter they live in, then we have an indicator of industriousness, discipline, responsibility, personal pride and of an aspiration to better one's life. These are all excellent predictors of honest and reliable borrower behavior.

Best practice in microcredit goes beyond just mastering these indicators of moral character. In addition, successful microfinance programs always try to **cultivate the intrinsic willingness to pay** among their target clientele. This can be done by financial education that explains the rights and obligations of borrowers, shows borrowers how to interpret the cost of credit and how to use a loan profitably. **Financial education** should also impress on clients the dangers of becoming entrapped in a debt spiral. It should make very clear what a responsible borrower may take on debt for and what expenses one should better not borrow for.

The role of credit reference databases

An important instrument for instilling borrower discipline is, of course, the threat of **being cut off from future access to finance**. This is where an independent credit reference database with broad national coverage comes in. The first reflex is always to interpret a credit bureau as a look-up tool on new borrowers that shows what they have already outstanding and how they have paid in the past. Yet, the credit bureau is at least as important as a means of enforcing discipline on existing borrowers. If we employ responsible collections principles, we cannot physically or verbally intimidate clients, we may not shame delinquent borrowers publicly, we may not terrorize them on the phone at all hours of the night etc. etc. The only legitimate and effective tool for shaking up a delinquent borrower is the threat of reporting their arrears to the credit bureau and thereby cutting them off from future credit anywhere and for a long time.

Operational risk in the credit process

And then there are always some cases, where a borrower is not just insufficiently disciplined to always put the loan installment first, but where the borrower - independent of ability to pay - will never pay and **never even had the intention to pay**. This is fraud, of course. It may happen spontaneously when an individual desperate for money will say whatever it takes to get a loan and then will never be seen again. It also happens in an organized fashion, where a criminal may send out runners, who will take on microloans and turn the proceeds over to the gang leader. It often also occurs with the collusion or even at the initiative of staff in MFIs. Losses incurred on loans made in the presence of fraud or willful disregard for policies and procedures of the institution are actually not credit risk in the strict sense of our definition. Rather, we will treat fraud and process failures in the lending process as manifestations of operational risk, which will be the subject of Unit 4.2.

Borrower Selection - Application Analysis

Practical **credit risk management must intervene right up front**, at time of *origination* (fancy speak for making a new loan), in terms of **prudent borrower selection**.

If you have had formal training in conventional bank credit operations, you will probably think of financial statement analysis, looking at credit bureau reports, studying external and internal ratings and scoring results etc. These would all be great tools for assessing willingness and ability to pay and thus making good borrower selection decisions.

Limited utility of financial statement analysis in MSME credit

The problem is that in MSME credit in developing & emerging markets, we generally have none or very few of these data inputs at our disposal. **Everything is informal.** There are no official credit records, or they are only now beginning to emerge. Financial statements where they are available are not credible or verifiable and often exist in multiple versions: an unprofitable version for the tax authorities, a strong one for the bank and an honest one for the owner, which you never get to see.

For these reasons, MSME lenders in developing markets have over the last 30 years refined **special lending methodologies and business rules for financing micro- and small enterprise** activities in informal environments. Most of you will be quite familiar with these types of processes and controls. They are the first line of defense in microcredit risk management. These processes serve all of the above goals simultaneously: They eliminate fraud (operational risk), they discourage impulsive borrowing behavior (e.g. financing a bachelor party on credit), they assess the true financial situation of the household and the ability to pay, and they determine the moral character of the client, i.e. his/her willingness to pay.

Here now is a short list of some of those verifications, controls and selection processes that MSME lenders have put in place in order to **manage credit risk at origination**:

Best practice controls prior to loan disbursement

- **Ask future borrowers to make regular savings installments first**, before borrowing. This demonstrates the existence of cash flow, discourages impulse borrowing and deters fraudsters.
- **Organize borrowers into solidarity groups** of people who know and trust each other enough to guarantee each other's loans.
- **Ask prospective borrowers to bring references, make reputation and character inquiries** with neighbors and community elders. This should weed out unreliable types, drunks and small time criminals.
- **Visit the home of the borrower**. A home visit is an excellent opportunity to check whether the life-style matches the financial situation disclosed, whether the number of dependents and additional adult contributors to the household budget is correct etc.

More best practice controls prior to loan disbursement

- **Inspect the business, activity or work place**, including a surprise visit. This would be the time to have the loan officer count the cash on hand and look at the records and cash book kept by the entrepreneur. It is good practice to recreate the actual cash flow for the last month by adding up invoices, counting stock etc.

- Review **credit bureau record** for the borrower and co-signors, where available. If the client is a **repeat borrower, check previous payment behavior** and incidence of arrears. Past lack of payment discipline is always a strong predictor of future trouble.
- **Make borrowers wait a few days or a week** before disbursement of an approved loan. If it is so urgent, then there often is a problem that the borrower has not told us about.
- **Require a second opinion** on a new borrower from a peer loan officer before disbursement.
- Apply **staggered approval limits by amount and by risk category** within the hierarchy from junior to senior loan officer, to branch manager, all the way up to head office credit supervisors. Often, the approval process is formalized into credit committees held at branch, regional and head office levels. An o**bjective risk categorization** could be obtained from an internal or external scoring or rating, for example.
- **Frequent random verifications** on loans approved within the delegated authority should be carried out by supervisors in the credit hierarchy.

The type of verifications and controls highlighted above have worked remarkably well in even the most difficult and informal MSME credit markets. This is because they acknowledge the reality that the issue of default is overwhelmingly a **question of *willingness* to pay under hardship**.

The *ability* to pay is always marginal in MSME credit. There is rarely a month in the life of a micro and small business borrower where each and every financial obligation has been met and significant money is left over. This is what being poor means. Even if one accumulates actual cash savings, then these are strictly an outcome of disciplined priorities. Savings are the result of not giving in to what the children or the husband want to spend the money on, but making real sacrifices in order to build up an emergency buffer.

All of this does not mean that the ability to pay is not important in making credit risk assessments. Absolutely, the financial situation of the borrower is a critical element. But it comes right after the willingness to pay. Just because the ex-ante ability to pay is typically already marginal, this does not mean that it could not get worse from there. **The risk then is that the ability to pay goes from marginal to impossible.** And when it is impossible, when the cash just is not there, then even the most willing and disciplined borrower will default.

Assessing the financial situation, the ability to pay

So, when **assessing ability to pay in microcredit**, we don't look for luxury problems, whether there is 3 times or 2.5 times cash flow to debt service coverage. We look at how volatile and vulnerable the modest coverage of cash-flow-to-installment is. It will always be tight, but is this household expense budget half-way realistic, is this monthly cash flow a rare, exceptionally good month, and is this an activity that is vulnerable to sudden changes in demand? Are there no alternative sources of income in the household that could tide the borrower over, should the main activity encounter a problem? Those are the issues that will push a micro-borrower over the edge and will make the situation go from scraping the installment together with iron discipline to giving up and defaulting.

Need for ongoing monitoring

Because the borrower situation in MSME credit is typically quite volatile, vulnerable to sudden reversals of fortune and intransparent, we cannot just lean back after loan disbursement.

A beginner in credit risk management might be tempted to say, what can we do after disbursement? The deed is done. Now, it is up to the borrower to survive in the tough microenterprise world and the lender must simply wait and see how it plays out. What does frantic running around and hyperactive monitoring change about the probability of default?

Monitoring: maintain willingness to pay

That's a simplistic probabilistic view of default. There is always **feedback from the monitoring activity to the actual client default behavior**. Monitoring serves two objectives: (1) maintaining the borrower's willingness to pay and (2) finding out early, if the financial situation of the borrower has deteriorated and is threatening the ability to pay. Most people would have immediately thought of the second objective, but the first one really is the more important rationale for monitoring. We must regularly monitor the client, and I mean not secretly from a distance, but actually make personal contact, visit the business, call the borrower's phone. We will find out how the business is going, whether there are emerging risk factors in the personal life etc. And most importantly, the client will know that we are still here and watching him like a hawk. The client must know that the loan officer will be personally disappointed, if a loan installment is late, that even a day's delay is a big deal and that severe consequences and additional costs will follow, if a loan falls into arrears. Monitoring is there to ensure that in the borrower's daily battle of the priorities under tight financial constraints, our loan installment always comes first.

Monitoring: early warning on deteriorating ability to pay

Coming back to the second objective of monitoring: what should an institution do, if a monitoring visit highlights a **deteriorating financial situation** in the borrower's business or signs of personal overspending and emerging **over-indebtedness**? Certainly, you should avoid throwing good money after now endangered money. Thus, no new disbursements to this client. This sounds obvious, but the line is not always easy to hold. What if this is a good repeat client who has hit a difficult moment? Giving just a little fresh liquidity may help the borrower right the ship and be forever a grateful good client. So, if you do go near a fresh disbursement in response to an increased default risk, you should do so only under the tightest scrutiny and with **additional good collateral**. That would be the time to ask the borrower to bring in the gold jewelry or hand over the car title or anything else that impresses the seriousness of the situation.

Monitoring: what to do when financial situation is deteriorating?

Another action variable following the detection of financial difficulty on an existing exposure, is to make accommodations on the existing loan schedule by stretching out payments over time or temporarily reducing the installment with an acceleration later. These kinds of **rescheduling actions** are also a dangerous animal. If we are too forthcoming with these types of adjustments, we destroy some of the carefully cultivated discipline and may give the client the impression he can walk all over us. Rescheduling must be tightly controlled within the credit hierarchy and rationed, so as to only benefit those cases, where we have optimal assurance that this is it, and that the new schedule will be faithfully adhered to.

Go easy on the rescheduling

You probably have heard of the shenanigans that some MFIs have played with rescheduling. When **rescheduling is handled loosely**, loan officers will be tempted to reschedule every loan in arrears and thereby reset the client to healthy current status. Obviously, a rescheduled loan should always be flagged as impaired or troubled and loan officers should not be collecting performing loans bonuses on rescheduled loans.

Also, in a **competitive situation** where multiple institutions might lend to your clients, we have to be careful not to weaken our position to collect via rescheduling. It could well be that by rescheduling and thus lowering the installment load on the borrower, we open up an affordability window (i.e. cash-flow-to-debt-service) for another lender. The competing lender could then squeeze one more new installment into the budget and max the client out again. We don't want to get into game theory here, but this could actually be a viable strategy for a lender who plays his cards right in a competitive lending environment:

1) Because of your superior monitoring process, you realize before any other lender that there is trouble brewing in the borrower's business.
2) You reschedule the loan by deferring a few installments in the near term that will be collected via an accelerated schedule later.
3) You subtly encourage the borrower to seek additional fresh liquidity from another lender, who will comply because the client is not in arrears on the credit bureau and the monthly installment load permits further borrowing.
4) Then you collect your loan essentially from the cash just borrowed elsewhere.

Yes, it's a dog eat dog world out there in MSME finance!

So much for now on the underlying nature of default in MSME credit. You were probably already wondering what is happening with all the borrower psychology, the anecdotes and all the words of wisdom on the MSME lending process. Where are the numbers, where are the hard statistics? Exactly, they are coming right up.

Chapter 3 – Review Questions

- Is default an ongoing process or is there a specific moment in time associated with default?
- Explain the nature of liquidity constraints faced by MSME borrowers.
- In how far and for which type of borrowers can an arrear be an indicator of credit risk rather than the manifestation of default?
- In how far are ability and willingness to pay inter-related in microcredit?
- If a borrower never had any intention to repay the loan, is this credit risk or operational risk?
- Name five elements of best practice pre-disbursement controls in microcredit.
- How can active borrower monitoring influence the probability of default in MSME credit?
- When is it appropriate to formally reschedule a loan into a new contractual installment plan?

4 Descriptive Portfolio Statistics

4.1 Arrears Basics

This Chapter 4 is all about analyzing arrears as a risk indicator.

We define an **amount in arrears,** or **an arrear** for short, as a currency amount that is due under a loan agreement but has not been settled when due. An arrear is specified by an amount and its age, typically expressed in days overdue.

If a borrower fails to make several scheduled payments, this gives rise to multiple parallel arrears, each with a particular age. If subsequently a payment is received from the borrower, it is applied towards settling the oldest arrear first, before relinquishing more recent unpaid amounts due.

A technical problem arises with this arrear definition in respect to current accounts with **revolving overdraft credit lines**, which generally do not have scheduled contractual payments. Here, lenders often begin counting arrear days on overdraft balances beginning with the 91st day without any credit received in the favor of the client.

Arrears as a credit risk indicator

Arrears can serve as an **indicator of deteriorating credit risk** and an increasing probability of default. This is true in consumer credit, in microfinance and in mass-market small business finance and anywhere else, where the loan amounts are relatively small, the borrowers numerous and subject to liquidity constraints, just like we discussed at length in the previous chapter. In corporate credit, where we would be dealing with businesses that have professional financial management and access to various lines of credit, this does not apply, of course. In the Enron case, arrears would not have been an early warning sign of risk, but a delayed confirmation of materialized default. Enron hit 90 days arrears on their loans outstanding exactly 90 days after filing for bankruptcy.

Arrears only work as a risk indicator in the presence of liquidity constraints

We are belaboring this distinction between corporate credit and small business finance in regard to liquidity constraints, because we are about to build a huge analytical edifice on the notion of arrears as an early risk indicator. So, we really want to make sure that we apply it to the right types of situations. Sometimes, in more developed markets, what we call SMEs can be rather large professionally managed family owned corporations. One should certainly avoid assessing credit risk in a German manufacturing SME with 500 employees by way of arrears behavior.

Figure 6 shows a sample **arrears aging schedule** that should be readily available from any debtor management system. It is a list of all loans outstanding with a breakout of arrear amounts and age. Often, the report is already reduced to only those accounts that display an arrear or is further subdivided by branch, by loan product, by responsible loan officer etc.

Arrears Aging Report as of [Date]				Amount in Arrears					
Client ID	Account No.	Client Name	Balance Outstanding	1-30 days	31-60 days	61-90 days	91-120 days	121-150 days	...
001	1234567	XXX	10,250.00	120	145	150	155	155	
002	2345678	XYZ	2,379.00	85	85	85			
003	3456789	ABC	4,569.00	115					
004	4567891	BCD	537.00	50	50				
...	...	...	...	...	...	...	...	...	...
Totals			**2,389,679.00**	**8,050.00**	**4,535.00**	**3,600.00**	**2,135.00**	**1,505.00**	**...**

Figure 6: Sample of an Arrears Aging Report

How to read an arrears aging report

The oldest arrear determines the overall arrears age of a particular loan. Thus, borrower XXX is 121+ days in arrears, which is certainly too far gone to talk about risk. This is a materialized default. Client XYZ is 61+ days and approaching default. Client ABC popped up for the first time on this report with a fresh arrear in the 1-30 days bracket. This is a newly emerging risk and needs immediate attention. We need to find out what the reason for the delay may have been and let the client know that the lender is paying attention and further arrears will not be taken lightly.

Even more interesting than the question who the individual borrowers are who are in arrears, is the portfolio arrears profile in the summation at the bottom. As these sums grow and spread to the right, or hopefully also sometimes decline and recede to the left, we can take the pulse of the risk trends across the entire portfolio.

Sometimes, the casual reader of an arrears aging report might misinterpret the relatively small arrears amounts relative to the total portfolio outstanding. One might think that only the particular installments not paid are now lost or at risk of being lost. That is obviously wrong: For client XXX the lender will likely have to provision or write off the entire amount outstanding. And even the first 1-30 days arrear posted for client ABC does not just put 115 dollars but the entire balance of 4,569 dollars into a higher risk category.

Portfolio at Risk

In order to avoid this common beginner's mistake in reading arrears reports, the MSME finance community has widely adopted the **Portfolio at Risk presentation of arrears**. Figure 7 gives the equivalent Portfolio at Risk (PAR) presentation for the sample arrears data in Figure 6.

Portfolio at Risk Report as of [Date]				Portfolio at Risk			
Client ID	Account No.	Client Name	Balance Outstanding	PAR1	PAR30	PAR60	PAR90
001	1234567	XXX	10,250	10,250	10,250	10,250	10,250
002	2345678	XYZ	2,379	2,379	2,379	2,379	
003	3456789	ABC	4,569	4,569			
004	4567891	BCD	537	537	537		
...	...	...	...	...	...	...	...
Totals			**2,389,679**	**215,071**	**143,381**	**71,690**	**59,742**

Figure 7: Sample Portfolio at Risk Report

PAR1 is defined as the total balance outstanding under any loan contract for which an arrear of age 1 day or higher exists. **PAR30** then is the total balance outstanding under any loan contract that displays an arrear of 30 or more days and so forth.

Again, PAR really **contains no additional information relative to an arrears aging report.** The only difference is that it highlights the gross amount of the exposure at risk of not being collected, now that some arrears have begun to emerge on certain loans. We should note that PAR is expressed in terms of **unimpaired gross exposure**, that is before deducting any potential loan loss provisions already accrued against these receivables.

How to interpret PAR

The most frequently cited PAR statistic in microcredit is the **PAR30** value. The level of 30 or more days in arrears is a reasonable **alarm threshold** where a microfinance institution should be seriously concerned about the potential full default of the borrower. If you moved the threshold much higher, say to PAR90, you are looking essentially no longer at a risk of default but at materialized defaults that should soon be moved over to the legal collections team. If you set the PAR threshold very low at just 1 or more days in arrears, the gross exposure at risk will tend to be inflated by many temporary arrears that may have been caused by "technical" issues such as due dates falling on a day just before or after a holiday weekend or bad weather keeping payers from coming to the branch to settle their installment etc.

Nonetheless, **PAR1** has its place as an early risk indicator, if considered not in isolation but together with the trends in PAR30 and PAR60. We never said that being one day in arrears constitutes a default. But taken across an entire portfolio, the fact that this month we have twice as much in loan balances impacted by arrears of 1-30 days, tells us that there is an emerging problem with payment discipline and/or a deteriorating borrower liquidity situation. They can still come up with the installment eventually, but it is becoming harder to scrape the funds together in order to pay on time, and more borrowers are missing the deadline. So clearly, a loan balance for which we register even a temporary delay of just a day or two is exposed to more default risk than a balance where the borrower pays on time every time.

In common MSME credit practice, the **PAR statistics are also analyzed in various segmentations**: by branch, by loan product, by borrower activity, by responsible loan officer etc. This is quite helpful in detecting risk and default concentrations in certain parts of the portfolio. Segmentation can also explain to some extent what the underlying risk factors are that may drive the arrears behavior. Many MFIs will use the results of PAR segmentations to argue points such as:

- Women borrowers are a lower credit risk than men.
- Retail trade in imported consumer goods is less risky than construction trades.
- Region 1 is riskier than Region 2.

- Repeat borrowers in credit cycle 2 and above are lower risk than first time borrowers.
- etc.

Don't over-interpret uni-dimensional PAR segmentations

However, let's be careful **not to overstretch the segmentation approach**. Uni-dimensional segmentations as in the examples above can indeed provide interesting indications for further analysis. Yet, it is rare that one would be able to identify a single risk factor that would by itself have so much explanatory power that we could really use it in decision making: for example, if Region 1 is really so risky, then we should not be lending there anymore, or only for much smaller amounts and with twice the collateral.

We have seen many microfinance practitioners **getting stuck in endless uni-dimensional PAR segmentations**. There are so many possible criteria and the differences in PAR rates between them are rarely statistically obvious and often contradictory. It also does not help to segment simultaneously by three criteria or more, say by looking at a table with PAR for women borrowers by branch and by activity. It just gets more confusing and the patterns that might exist are so subtle that it would be hard to pick them up reliably with the naked eye by just reading a segmentation table.

Statistical scoring to the rescue!

That is too bad, because the idea of studying arrears and default rates under multi-dimensional simultaneous segmentation actually is the way to go in order to make sense of the available data. So, instead of simply saying men have a slightly higher average default rate than women, we have to ask the question differently: What about married men in Activity 2, in Region 3, with 5-10 years of experience, with 2-5 employees, on their third loan with us, with an installment to cash flow ratio of 25%, who own their home and have a fixed line telephone at their place of business?

These type of multi-layered questions we certainly cannot answer across thousands of loans just by looking at PAR tables. But that is what modern statistics were invented for. A statistical credit scoring model will give you exactly the optimal combination and weighting of those segmentation criteria or risk factors, which will best set apart those borrowers that will likely default from those who will be good payers. And it's not difficult. We will show you step by step how you can do this at home with standard Excel tools in Chapter 7.

First, however, let's see what other useful statistics we can squeeze out of the simple idea that arrears are an indicator of emerging economic difficulty and therefore a predictor of default.

4.2 Vintage Curves

Purpose and Interpretation

Vintage curves are an important **arrears-based portfolio statistic** that can provide valuable early indications of default trends across the entire portfolio or certain segmentations of the portfolio by product, region or business units.

Vintage curve definition

A vintage curve is a visual presentation of defaulted balances by time elapsed since loan disbursement. The bad rate in this calculation is defined as the balances outstanding that are impacted by arrears of a certain severity divided by the original amount disbursed under this portfolio. Typically, the portfolio is clustered by month of origination such that **each production month becomes a distinct sub-portfolio** and a separate line in the vintage curve graph.

Figure 8 shows an example of vintage curves for a micro-consumer loan portfolio at African Bank, South Africa. Each colored line represents a monthly vintage (meaning "harvest") of loans originated during that month.

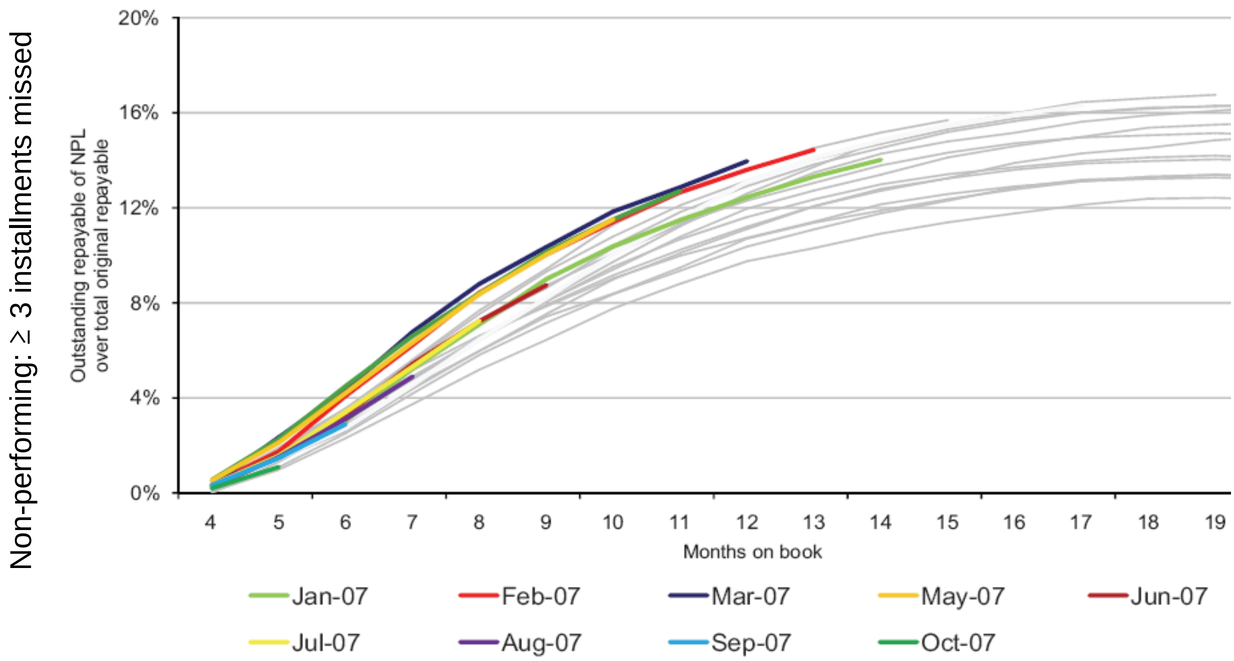


Figure 8: Example of Micro-Consumer Vintage Curves at African Bank, South Africa

Vintage curve interpretation

The vintage curves for African Bank are particularly interesting because they narrowly conform to a well controlled pattern and top out in the range of 12%-16% bad rates. This would be typical of a mature lending franchise that has exceptionally good visibility of credit risk, as it can only be achieved with well designed statistical scoring or rating systems. It should be noted that because of its high-price / high-risk micro-consumer business model, African Bank could operate very profitably on default rates of 15%-20%.[2] On closer inspection, one will notice that the more recent originations in Figure 8 actually converge towards a lower bad rate of 12% rather than the previous benchmark of 16%. This is the result of a conscious tightening of the scoring model in response to the introduction of consumer protection rules effective June 2007 that limit allowable interest rates and other charges.

Vintage curves are a leading indicator of portfolio performance

The significance of **vintage curves as an early warning indicator** becomes obvious, if one imagines that one of the more recent curves deviated significantly from the expected grey trajectory. A monthly group of loans that in Figure 8 had already reached 12% bad rate after 6 months would immediately trigger an investigation into the sources of the unusually high delinquency. In a conventional portfolio-at-risk statistic, however, this significant deterioration of performance would initially disappear in the global portfolio average. It would only become visible much later, when the unfavorable developments behind this trend have spread much further into the portfolio base.

We cannot emphasize enough this important advantage of a vintage curve analysis relative to traditional PAR statistics. **PAR is a lagging indicator** of portfolio performance, while **vintage curves can give early trend signals**. This is true both in an emerging portfolio crisis as well as when portfolio performance is improving again after a non-performing loans episode.

The **fundamental problem with PAR** is that it does not explicitly take into account that default is a function of time. Recently disbursed loans always display lower default rates than older generations of loans. How can installment loans that were disbursed two months ago be in default, if the default definition is to be more than three months in arrears? And even four months into the life of an installment loan, only those loans would show up in PAR90 that never paid a single installment. Even the worst borrowers generally manage to pay the first monthly installment with the liquidity that they just borrowed. So clearly, it is a law of physics that **new loans will always display lower bad rates than older loan generations**.

The problems with PAR

Therefore, if a portfolio is rapidly growing, there will always be more new loans than older loans and the defaults on the older loans will be diluted by the all the new loans that have not had a chance to default, yet. This is the lagging effect of PAR. How many times have we seen this same scenario play out: the portfolio is growing exponentially, PAR is a fraction of a percent, bonuses are paid and rock stars of microfinance are born. And when the portfolio growth slows, as it must eventually, all of a sudden PAR rises, or so it seems. In reality, PAR only converges towards the underlying default rate for seasoned loans

[2] We should note that African Bank did eventually collapse in 2014 and had to be placed under curatorship by the Reserve Bank of South Africa. What may have pushed African Bank over the edge was the massive support for the retailer Ellerines by way of in-store credit sales to low income households. Ellerines is part of the ABIL group that also owned African Bank.

that are old enough to have had a chance to default. In reaction to such an emerging crisis, many MFIs will reduce new disbursements and the portfolio begins to shrink. Pretty soon you have a majority of old loans on your books and the older they are, the higher the percentage that are in default. If you stopped new disbursements all together, you could have PAR converging to 100%, as the non-defaulted loans are paid-off and the remaining portfolio consists exclusively of loans in arrears.

All of this **confusion with PAR could have been prevented by looking at vintage curves**, where we clearly see the bad rate as a function of time since disbursement. We also don't have a size bias in comparing defaults at a certain age point across multiple years of operation, because the bad rate is expressed as a percentage relative to originally disbursed amounts.

Figure 9 is an example of vintage curve trends at a leading MFI in Morocco. This MFI went through exactly the kind of cycle we described above: rapid portfolio growth with low PAR -> growth slows -> PAR crisis -> portfolio shrinks -> PAR gets worse -> new tightly controlled loans perform better. But nobody believes it, because the lagging PAR is still high.

In this case from Morocco, 2008 was the year the PAR crisis broke through. We grouped all of the 2008 monthly vintages together and show the average bad rates of those 12 vintages at each age point in one blue curve. The bad criterion for the analysis in Figure 9 is 90+ days in arrears. We see that the loans disbursed in the 2008 crisis year topped out with a bad rate of about 9% after 24 months. Loans disbursed in 2009 and in the first quarter of 2010 performed still worse, reaching almost 12% terminal bad rate. But since then, the vintage curves highlight a consistent normalization of the bad rate.

The colored lines are materialized observations, while the grey continuation lines are forecasts. Shorter colored lines represent younger quarterly or monthly loan originations. The positive trend now is immediately obvious in that shorter colored lines almost always stay below the longer lines at each age point. With the trend continuation line, we can see that the forecast for the terminal bad rates on recent generations are only about 5% and falling.

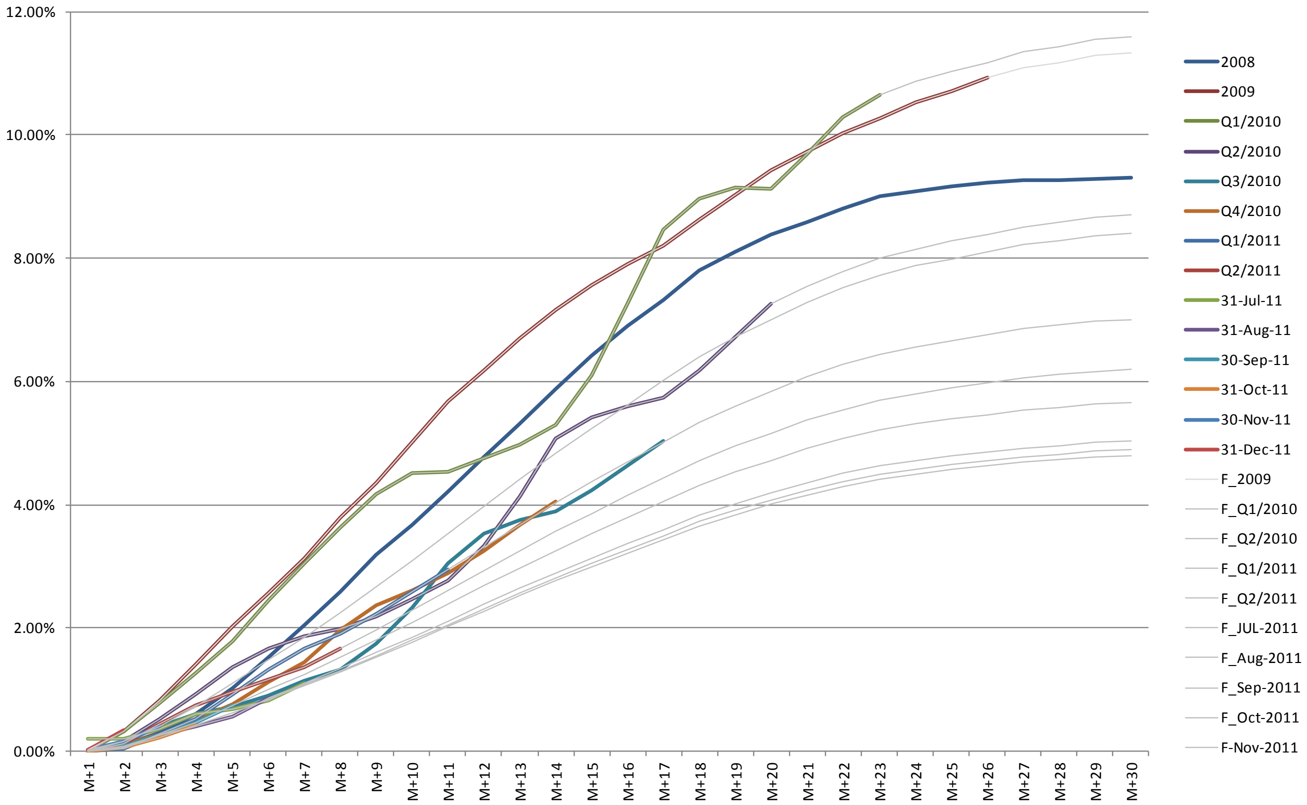


Figure 9: Post Crisis Vintage Curves at Moroccan MFI

And just to be sure it's crystal clear, we are not making the same mistake here now as before with PAR. Of course, newer loans perform better than old loans, but the vintage curves prove in this case that these newer loans perform even better than the old loans when they were young, too.

Discussing Figure 9 with the management and funders of the Moroccan MFI was one of the finest moments in vintage curve history. At that time, the global PAR90 was still in the unacceptably high 9% range. And since these 9% mostly were hardened bad loans from 2008/2009 that never were going to be collected, the only way to bring the PAR down in the absence of portfolio growth, would have been to write them off and thus eliminate the bad loans from both the numerator and denominator of the PAR ratio. But when management did that, funders would say, no-no, you are just trying to manipulate the PAR down, you have to add the written off loans back into the PAR ratio. In the meantime, the new loans made since mid-2010 really did reflect much improved borrower selection and internal controls, but everyone was still punishing management for the PAR from 2008 and 2009. The MFI was stuck in a mission impossible: we won't allow you to grow the book again until PAR (including the historic write-offs) comes down. Then vintage curves came to the rescue, and we could prove that the new book was better and that the improved disbursement processes deserved to be scaled up with additional funding.

Calculation and Data Requirements

In order to calculate vintage curves, the following **essential data elements** from the debtor management system are required:

- Loan ID
- Observation Date
- Disbursement Date
- Disbursed Amount
- Outstanding Principal on Observation Date
- Arrears Status on Observation Date (Good/Bad or arrears in days)
- Counter: number of months between Observation Date and Disbursement Date
- Additional classification elements: type of loan, branch, etc.

Generally one would use at least **24 months of consecutive monthly observation dates** to obtain a meaningful vintage curve window. It is important that the definition of the impairment ratio or the bad rate properly takes account of written-off loans:

Bad Rate = (Written-off Principal + Impaired Principal) / Disbursed Amount.

Unless a vintage curve model is already implemented in the core banking software, the curves are generally produced based on a database extract with those minimum data fields as above for multiple observation dates.

Vintage curve calculations

One then uses a **select query to manage the filtering of observations** from this global database: the filter criteria are the market segment or product (individuals, micro, SME) and the origination period. For each product and origination date filter setting, we need to retrieve the **original disbursed amounts** and the impaired amount outstanding for different age counter values. **A pivot table report** is used to group and sum the observations of “impaired outstanding” and “original disbursed” by the value of the counter, i.e. the number of months elapsed since disbursement. The actual vintage curve then is the quotient of the outstanding impaired amount (OutstandingLEK) **divided by the originally disbursed amounts** (ApprovedLEK). See the example of a calculation at an Albanian bank in the screenshots below.

Microsoft Access - [VintageImpaired

Counter	Sum of ApprovedLEK
0	1371332435
1	1771722898
2	1982346679
3	2328252733
4	2187960369
5	2081906158
6	1962007090
7	1741814119
8	1287396146
9	1439151332
10	787593162.6
11	706090590.4
12	574299458.1
13	337226189
14	240272581.1
15	189937836
16	10000000
Grand Total	20999309777

Counter	Sum of OutstandingLEK
0	75900000
1	24086866.79
2	98149411.05
3	99597884.14
4	119646968.6
5	162935418.6
6	93476115.6
7	37928691.66
8	148622907.5
9	173882496.6
10	13665227.8
11	87026294.69
12	81404832.32
Grand Total	1216323115

Figure 10: Pivot Table Reports for the Vintage Curve Calculation

In principle, this can all be done also in Excel using lookup functions and pivot tables, but the data file generally becomes too large to handle in Excel with memory-intensive lookups. This is mainly because we need a full inventory of all loans made over the course of three years or so plus a monthly list of all loans in arrears at a minimum of 24 observation dates. That quickly puts us into hundreds of thousands of data lines in Excel.

In order to let you go through the calculation steps in Excel nonetheless, we produced a simple portfolio using randomly generated defaults and loan amounts with just the barebones data elements needed for the vintage curve calculation. See the Excel file M4.1_Ex5_VintageExercise.xlsx and the partial screen shot below.

			------------- Loan in 90+ days arrears on Observation Date? Yes=1 No =0 ---											------------- Loan Balance Outstanding on Observation Date ------						
LoanID	DisburseDate	DisbursedAmount	30-Jun-10	31-Jul-10	31-Aug-10	30-Sep-10	31-Oct-10	30-Nov-10	31-Dec-10	31-Jan-11	28-Feb-11	31-Mar-11	30-Apr-11	31-Mar-09	31-Oct-09	31-Mar-10	30-Apr-10	31-May-10	30-Jun-10	31-Jul-10
Loan1	31-Jan-09	1,700	0	0	0	0	0	0	0	0	0	0	0	1,540	926	432	327	220	111	0
Loan2	31-Jan-09	2,400	0	0	0	0	0	0	0	0	0	0	0	2,174	1,307	610	462	311	157	0
Loan3	31-Jan-09	3,700	0	0	0	0	0	0	0	0	0	0	0	3,351	2,014	940	712	479	242	0

Figure 11: Partial Screenshot from M4.1_Ex5_VintageExercise.xlsx

Exercise M4.1_Ex5: Vintage curves

There are 9,000 loans disbursed over 36 months, each monthly vintage consisting of exactly 250 loans. The loan amounts vary randomly between 1,000 and 5,000 and are all at 24% p.a. over 18 months with fixed monthly installments. The file provides you with the essential data needed for the vintage curve analysis: LoanID, Disbursement Date, Disbursed Amount plus the default/nodefault at 36 consecutive monthly observation dates (red titled columns). Further to the right, we have assembled the balances outstanding for each loan at each observation date (light blue titled columns). These balances are equal to the principal balances under the 18 months loan schedule unless the loan is in default from which time onward the balance does not decline further.

Now it's your turn to work with the data in M4.1_Ex5_Vintage Exercise.xlsx, in order to produce the vintage curve graph. The necessary information is all there, you just need to do a few smart manipulations to make the numbers line up nicely for a chart. The goal is to have a table that shows for each monthly generation of loans how much defaulted balance (90+ days in arrears) was outstanding at each age point. Good luck!

The final graph should look something like this:

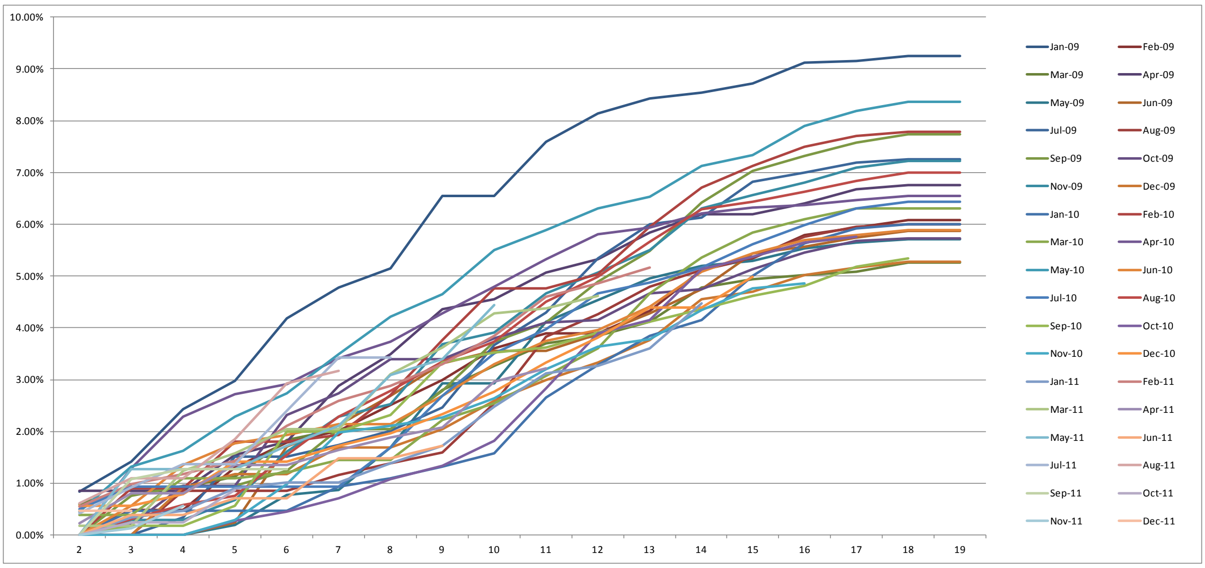


Figure 12: Solution for M4.1_Ex5_VintageExervise.xlsx

M4.1_Ex5: Solution

A separate file with the solution is provided in M4.1_Ex5_Solution.xlsx. In cell CA2 in sheet VintageRawData, we still left the formula that we copied into the entire range CA2:DK9001 in order to get the defaulted balance by age in the black titled columns. The pivot table report in the sheet Pivot table is still "live" so that you can inspect how we arranged the fields for the report. We then just copied the values from this pivot report into the sheet VintageCurves, where we obtained the percentage values of bad balances relative to original disbursed and produced the chart.

Usually, raw portfolio data for vintage curve analysis does not come out of the system as neatly arranged as in the previous example M4.1_Ex5 with one line per loan and all the observation dates set up in columns. You can typically expect to get identically structured monthly portfolio inventories with arrears status and balance outstanding for one particular month end, one such inventory per Excel sheet. Instead of setting the consolidated sheet up 9,000 loans * 36 monthly observation columns as in M4.1_Ex5, you would just append the monthly tables below each other. This would create a master table that is 9,000 * 36 monthly observations long and repeats the observation date and the static loan information at the front of each row. You would typically still have to insert a column with an age counter into this master table:

=ROUND((ObservationDate-DisbursementDate)/30,0).

The trick in getting to the vintage data from there is to insert a pivot table which groups the rows by disbursement month (i.e. the vintage), arranges the columns by age counter and puts the sum of the balances outstanding inside the pivot table area. Then you set the filter to default=1 et voilà: you have for each monthly vintage the ramp-up of defaulted balances outstanding by age.

Practical calculation challenges with vintage curves

However, this usual process will only work in Excel 2007 or later where you can have more than 64,0000 lines and with a really powerful processor and lots of memory, otherwise the pivot table will crash. Therefore, we normally would import and append the monthly observation data sheets directly into a MSAccess database table. There you have an identical Pivot Table report function, that gives you the same output. It just won't crash because it works sequentially and does not hold all the queries in memory like in Excel. For those of you who know MSAccess, we also included an example of a vintage curve calculation along those lines with the course files: M4.1_Ex6VintageExample.mdb.

Vintage Curve Extension Forecasts

How to derive a vintage curve forecast

Next, we will show you how we produced these really helpful grey continuation lines in Figures 8 and 9. The idea of these lines again is to provide a forecast of how the bad rate for a particular vintage might further evolve towards a terminal level reached towards the average scheduled final maturity of those loans. This can be done by applying an **assumption about a typical shape of vintage curves** beyond the already materialized bad rates.

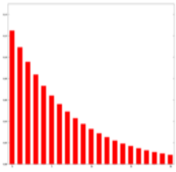
It appears that vintage curves always grow steeply the first few months, then go through an inflexion point, gradually flatten out and asymptotically converge towards a terminal level. Sometimes, curves fall back again slightly, once the age exceeds the scheduled maturity of the vintage. At that point, really only the defaulted loans are left in the vintage. If through legal collections some principal eventually gets recovered, the bad rate could taper off. However, measuring effectiveness of legal collections is rarely the objective of a vintage analysis. We therefore often cut off the presentation of the curves at the age that corresponds to the average maturity of the portfolio.

So, how do we express a standard shape expectation mathematically? We opted against trying to fit a specific exogenous mathematical function to the behavior of the curves. Instead, we extract the standard shape from our own recent vintage curve history. For this, we simply divide a materialized bad rate at age x by the bad rate at age x-1 to get a growth rate multiplier for the transition from bad rate x-1 to age x. We average these growth rates at all age transitions for which we have materialized bad rates in the database. From the last bad rate observation for a particular vintage onwards, we extend the grey line simply by multiplying the bad rate with the series of month-by-month growth rates. Easy enough.

In order to make it a bit more fancy, we decided to not just take an arithmetical average of the observed growth multipliers by age, but instead use an exponential moving average that gives more weight to recent materialized growth rates than to older ones.

An **exponential moving average (EMA)** applies weighting factors which decrease exponentially, thus giving much more importance to recent observations while still not discarding older observations entirely. The degree of weighting is expressed as a constant smoothing factor α, between 0 and 1.

$$EMA = \frac{p_1 + (1-\alpha)p_2 + (1-\alpha)^2 p_3 + (1-\alpha)^3 p_4 + \cdots..}{1 + (1-\alpha) + (1-\alpha)_2 + (1-\alpha)^3 + \cdots}$$



The weights in the average decline rapidly along a smooth exponential curve as in the picture to the right.

Using the vintage curve forecast template

Now, take a look at the model file **M4.1_Ex7_VintageCurveForecast.xlsx**. This is where we derive the standard growth rate shape and produce the vintage curves. The lightly green colored zone in the sheet DataCapture_AverageShape is the input area to where you can copy/paste or link your percentage curves obtained from an external calculation. To the right of this input area in the same sheet, we calculate the average growth rate curve. The final averaged result is in line BA66:CK66. The second sheet in this workbook FinalCurveswithForecast regroups the percentage curves for display and attaches the actual graph. Regrouping all three vintages of an older quarter and displaying their average bad rates per age point, as we did here, is helpful in reducing the number of individual lines on the chart. It is ok to show all curves in a mass display like in Figure 8, if we simply want to prove the point that all curves are huddling close together and nothing is really sticking out from the usual pattern. Yet, if you want to interpret individual vintage curves, it is generally helpful to cluster older vintages into a single quarterly or even annual average curve.

Common Calculation Errors and Issues

We hardly ever get meaningful vintage curves on the first calculation attempt. Here are a few of the typical issues we might encounter.

Don't divide by portfolio currently outstanding, divide by original disbursed amount

Some institutions like to define the bad rate in the vintage curves in close analogy to portfolio at risk. This would mean we divide the principal outstanding which is in the particular level of arrears that defines "bad" by the total portfolio outstanding from that vintage, instead of dividing by the total amount initially disbursed. This figure is always distorted because the denominator shrinks with the principal amounts that are being paid back every month. It becomes completely meaningless the older the vintage gets, as more good principal is being paid back and we are essentially left with bad principal in the numerator and the same bad principal in the denominator. The bad rate hence converges to 100%.

Include write-offs in the bad rate

Another technical problem may arise from the **treatment of write-offs**. As loans are written-off, they are generally moved to a separate legal debt management system and no longer appear in the standard monthly loan inventory and arrears aging report. This may give the impression that the bad rate in a certain vintage has suddenly recovered massively. A legitimate drop in the bad rate is certainly possible. It would require that more bad loans by value start paying down their arrears and become good again than other loans miss installments and join the bad statistics. This is quite rare, particularly, if we put the bad threshold at a relatively high level of arrears such as 90+ days. Most often then, a drop in the bad rate is a calculation error. We simply forgot to add back into the numerator the written-off principal from for the particular vintage. **Once written-off, forever bad** is the rule. The same rule should be applied to loans that are rescheduled in the presence of already manifested arrears. The restructured principal should be forever retained in the bad rate of that vintage.

Treat rescheduled balances like write-offs

Another common reason why we might get a spaghetti effect rather than a nice bouquet of vintage curves is **setting the "bad" arrears threshold too low**. If one sets the threshold to an extremely sensitive level of just 1d+ in arrears, one would certainly get rather volatile curves as clients transition in and out of the default level depending on minor variations in payment discipline. There is a better tool for studying these transitions between early stage arrears than a vintage curve chart. This would be a transition matrix, which we will study next, see Chapter 4.3.

No micro-segmenting, please

When **segmenting vintage curve data** by product, region or type of client, one has to make sure that the remaining vintages are granular enough to produce stable average behaviors. If there is only a handful of large SME loans in a vintage, the bad rate might easily become totally erratic, i.e. pushed up or down by the payment behavior of a single borrower. That's an obvious problem. Statistics only work in the presence of a large number of atomistic cases.

Finally, here is kind of a luxury problem with vintage curves. The vintage curves in Figure 13 are from a large MFI in Cambodia. It looks like another spaghetti situation.

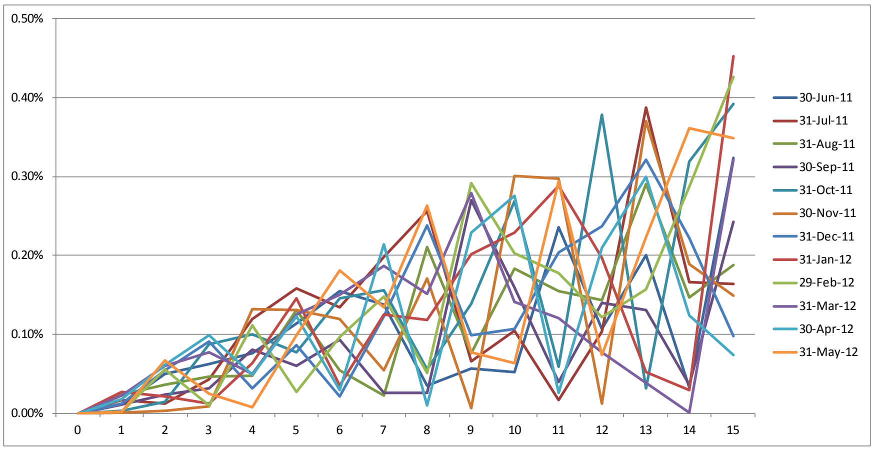


Figure 13: Example of a Vintage Analysis at a Cambodian MFI. Zoomed to 0.1% bad rate

Don't zoom in too far

The MFI actually tried to interpret these charts and has speculated about reasons why particular vintages were displaying certain peaks and drops etc. In fact, there is nothing to discuss here. You should simply not have zoomed in so far into the graph by setting the y-axis to 10ths of a percent bad rate.

If we **zoom out to a reasonable level** of full percentage points bad rate, the result is a flat line, as in Figure 14. That is a nice problem to have. Congratulations, the bad rates are consistently low and the new vintages are as good as older ones. Yet, even for this enviable MFI, vintage curves are not a wasted effort. They can still serve as an **early warning system**. Figure 14 is like a blank radar screen. Just because there is no danger approaching at this time does not mean that the radar is worthless. If one day, there should be trouble brewing in the portfolio at this MFI, it will manifest itself in rapidly escalating bad rates on new vintages and it will show up on the vintage radar, long before it impacts the average PAR across the entire portfolio.

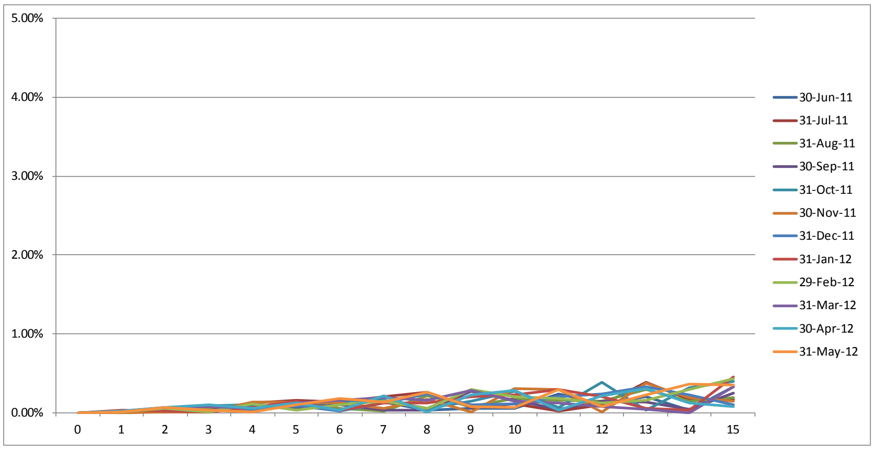


Figure 14: Same Vintage Analysis as in Figure 8 with proper Scaling of the y-Axis

System Implementation

Automate the vintage curve extraction with a script

We apologize if you are getting the feeling that this was maybe a little too much detail on the finer points of vintage calculations. We hope that you agree nonetheless that vintage curves are a **powerful way to visualize loan portfolio performance**. We put you through all the excel mechanics, because if you cannot produce a prototype calculation from your institution's existing data, then you will never have vintage curves. Of course, the risk manager should not spend days every month trying to copy/paste together a vintage report. Your **IT team should program a script** that will produce the formatted data for curves as part of the month end processing. The type of excel calculations that we showed here are simply the proof-of-concept or prototype that will document the data sources and the logic of their manipulation. The best way to explain to an IT person what a vintage report should do is always with a functioning prototype that spells out the logic in excel formulas rather than writing a long specification. Building an excel prototype analysis first, will also help with quality control and auditing of the system. If you can reproduce the same results on foot with your prototype calculation, then you know the application is working properly.