Contents

Chapter 0

Preface

0.1 About this manual

First and foremost, this is a collection of thoughts collected to write down and work through in order to be able to manage and analyze work at **AVANT** more efficiently. It appears non-trivial to understand and structure consumer lending as a business especially for people with no degree in finance or business with no dedicated learning materials.

For the like backgrounds in academia close to numbers, an analytical formula is preferable to any sequential or iterative problem solving whenever possible. In particular, this approach seems useful in learning the basic mechanics of the underlying processes. It is easier to expand into daily operations from a solid base of knowledge regarding the product.

This manual was started in January 2015 and has been expanded and improved ever since.

0.2 How to use this manual

This manual is structured into several different sections, not all of which are relevant to all readers. All parts of the manual marked with a colored bar on the side of the text are not essential, though instructive and can be skipped over. There are specific environments for examples, side notes, limits & approximations, consistency checks and references.

This manual tries to derive key facts from most basic assumptions whenever possible. Exact solutions are preferred, even though, hardly ever they are necessary or possible to analytically formulate.

Example(s):
Examples are used to give gray theory some color and further understanding.

Consistency Check:
Consistency checks reconcile counter-intuitive parts of the computation with known functional dependencies.

Side Note:
Side Notes point to noteworthy facts or mark short-cuts taken and simplifications applied.

Limits & Approximations:
Limits, approximations and simplifications for corner cases were worked out whenever feasible to provide the reader with an enhanced understanding of implications of the type:

- ***to achieve effect X, manipulate property Y in fashion Z*** or
- ***property A has functional dependence B but approaches easy-to-handle approximation C in case D***.

Reference & Literature:
References are given to give credit when credit is due to other publications.

Since *a picture is worth a thousand words*, illustrations are used to communicate key dependencies whenever possible.

Disclaimer: This manual does not claim ultimate knowledge, truth, completeness or authority on lending, credit reporting, modeling or the direction taken at AVANT. It has not been sanctioned or legally reviewed by any authority at AVANT.

I sincerely hope this manual is as helpful to you as it continues to be for myself. I encourage each and every reader to contact me for improvements, corrections and thoughts. Thanks for reading,

Stefan Hansel,
Chicago, 2018/01/05
Email: stefan.hansel@avant.com

Chapter 1

Introduction

1.1 Avant

Avant was founded in late 2012 by Al Goldstein, John Sun and Paul Zhang to provide a credit solution for the American middle class that has been underserved by the big banks after the great recession of the late 2000's. The company mission is:

Lower the costs and barriers of borrowing.

Primarily focused on the American middle class, customers have so called near prime credit scores [1].

1.1.1 History

Incorporation, first loan, average rate, figure of employee count, loans count, issued amount over time, expansion UK, 1B day, anecdotes from Adam's email, move to new location, addition of new offices, Philippines call center, Knoxville call center, VSP, Profitability, new product launches, cash flow positive

[1] Or rather *average* credit scores, in FICO scores this represents a range of roughly 580-720.

Table 1.1: Funding Rounds, Investors and overall funding

Round	Date	Amount	Investors
Series A	Q2 2013	$9M	August Capital, Victory Park Capital
Series B	Q3 2013	$25M	August Capital, QED Investors, Victory Park Capital
Series C	Q1 2014	$75M	Tiger Global Management, August Capital, QED Investors
Series D	Q4 2014	$225M	Tiger Global Management, August Capital, DFJ Growth, RRE Ventures, Peter Thiel, KKR & Co.
Series E	Q3 2015	$325M	General Atlantic, August Capital, DFJ Growth, JP Morgan Chase & Co., Tiger Global Management, Balyasny Asset Management, RRE Ventures, Hyde Park Venture Partners

1.1.2 Key People

group this, link to their linkedins, join dates, short bio

Founders:

- Albert Goldstein, Chief Executive Officer
- John Sun, UK Managing Director
- Paul Zhang, Chief Architect

Other Executives:

- Adam Hughes, President & Chief Operating Officer, 1st employee
- Robert Reynolds, Chief Technology Officer
- Suketu Shah, Chief Financial Officer
- Ryan McLennan, General Counsel
- Stacee Hasenbalg, Chief Compliance Officer
- James Paris, EVP Capital Markets and Corporate Development
- Grant Miles, Chief Risk Officer
- Chris Armsey, Senior VP of People
- Shyama Rose, Chief Information & Security Officer
- Kevin Lewis, VP of Business Development
- Bhanu Arora, VP of Analytics
- Charles Whittaker, VP of Product
- Matt Bochenek, VP of Operations
- Campbell Gibson, VP of Public Policy
- Anna Fridman (*)
- Jeffrey O'Dell (*)

1.1.3 Organization / Departments

Table 1.2: Consumer Credit, Gross Domestic Product for the largest economies by consumer credit. Percentage values of global share of credit and GDP as well as credit of GDP for 2016. Sources: FinAccord, World Bank for 2016.

	$B Volume[2]		% of world wide		% Credit	Population	$ per capita	
Economy	Credit[3]	GDP	Credit	GDP	of GDP	M	Credit	GDP
United States	18,056	18,569	40.8	24.5	97.2	322.2	56,046	57,638
China	3,771	11,199	8.5	14.8	33.7	1,378.7	2,735	8,123
United Kingdom	2,044	2,619	4.6	3.5	78.1	64.9	31,490	40,341
Germany	1,668	3,467	3.8	4.6	48.1	82.4	20,242	42,070
Canada	1,513	1,530	3.4	2.0	98.9	36.3	41,686	42,158
Japan	1,504	4,939	3.4	6.5	30.4	127.0	11,843	38,901
France	1,392	2,465	3.1	3.3	56.5	66.9	20,807	36,855
Korea, Republic	1,276	1,411	2.9	1.9	90.4	51.2	24,904	27,539
Australia	1,209	1,205	2.7	1.6	100.4	24.1	50,114	49,928
Switzerland	759	660	1.7	0.9	115.0	8.3	91,884	79,891
Netherlands	754	771	1.7	1.0	97.9	16.9	44,691	45,670
Spain	722	1,232	1.6	1.6	58.6	46.3	15,608	26,640
Italy	691	1,850	1.6	2.4	37.3	60.3	11,450	30,675
Rest of world	8,950	23,724	20.2	31.4	37.7	5,137.0	1,742	3,196
Totals	44,310	75,642	100.0	100.0	58.6	7,422.4	5,970	10,191

1.2 Consumer Credit 101

The expansion of Consumer Credit is one of the great economic achievements of the past century.

Virginia Postrel, The Case for Debt, The Atlantic, Nov. 2008

The total addressable market of consumer credit in the US and worldwide is a pitch slide you see flying around a lot. Typically, a number of 10 trillion USD for the overall market size in the US is used. It is a nice round number based on the overall outstanding balances of consumer credit including residential mortgages and certainly the right order of magnitude - or is it? It is worthwhile to understand the origin and implications of this number. Some questions other than the pure value of this number are of interest:

- Is outstanding balance the right way of determining or characterizing the addressable market?
- Or should that be a rate of issuance per quarter?
- Is the data source representative?
- What about the consumers currently not served?

Table 1.2 gives an estimate on the size of the consumer credit market and gross domestic product for the leading world economies. The data is based on the World Bank and FinAccord publications. Overall, consumer outstanding balances correlate well with GDP. The global volume of consumer credit in this statistic is around $44T at $76T GDP. The United States account for over 40% of global consumer credit volume at

[2]Based on publications of the main regulator / statistical body of the respective countries, converted to USD by the average mid point rate of OANDA.com

[3]Based on balances outstanding; reported credit is lending to private individuals and non-profits, focus on residential mortgages and consumption.

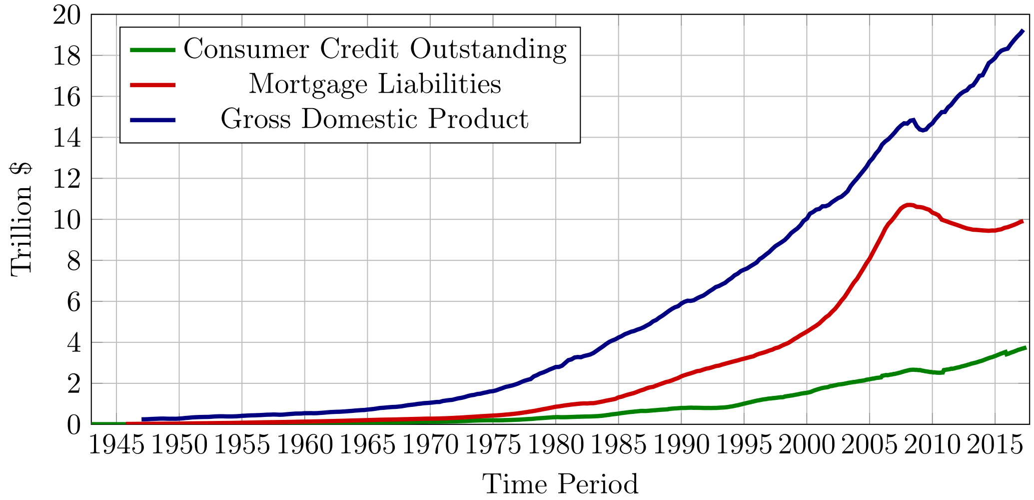


Figure 1.1: Outstanding balance on consumer credit, owned and securitized, mortgage liabilities of households and non-profits, gross domestic product according to Federal Reserve Economic Data. Details in the text (see section 1.2.1).

≈ 24% of global GDP. The US has the second highest per capita credit obligations after Switzerland but is much larger as an economy where it clearly ranks first. We will focus on the US.

The Federal Reserve Economic Data given in figure 1.1 shows the development over time of all consumer balances outstanding excluding those secured on real estate, mortgage liabilities for households and non-profits[4] as well as the gross domestic product. Mortgage balances are the bulk of outstanding credit held by consumers. Combined the number approaches $14T , or 75% of GDP. Consumer credit and GDP grow in lockstep, if they don't that is an indicator of a lack of macroeconomic balance.

Figure 1.2 shows data published by the Federal Reserve of New York's Quarterly Report on Household Debt and Credit and the TransUnion Industry Insights Report. Both data sources generally agree[5] and put the outstanding balances on residential mortgages and home equity lines of credit in the order of $9 trillion and other credit in the order of $4 trillion, with another $3 trillion open to draw on the revolving lines. So based on this data, the market size in the United States is $13 trillion with another $3 trillion open to draw. However, the credit products above work on very different time scales and magnitudes of credit, e.g. mortgages work on a decades long time scale in the 100's of thousands of $, where credit cards work in the few thousand $ balances in a fast paced revolving environment. The net increase in balances by quarter is $220B overall, driven mostly by mortgage balance growth. So a more detailed look provides additional insights.

While the balances will continue to grow in the future with income, home values, inflation etc., overall demand is limited. There are only so many homes to be financed, so many cars to be bought by the consumer that a clear idea of the number and needs of the consumers will provide realistic assessments of the opportunity at hand. The number

[4] The exact definitions of consumer credit are given in section 1.2.1

[5] Mortgages, HELOCs and Cars coincide between the two, Credit Cards of the Fed are split up into bankcards and private label for TU, then match. The type *Other* of the Fed entails among others the personal loan balance which is specifically broken out in TU.

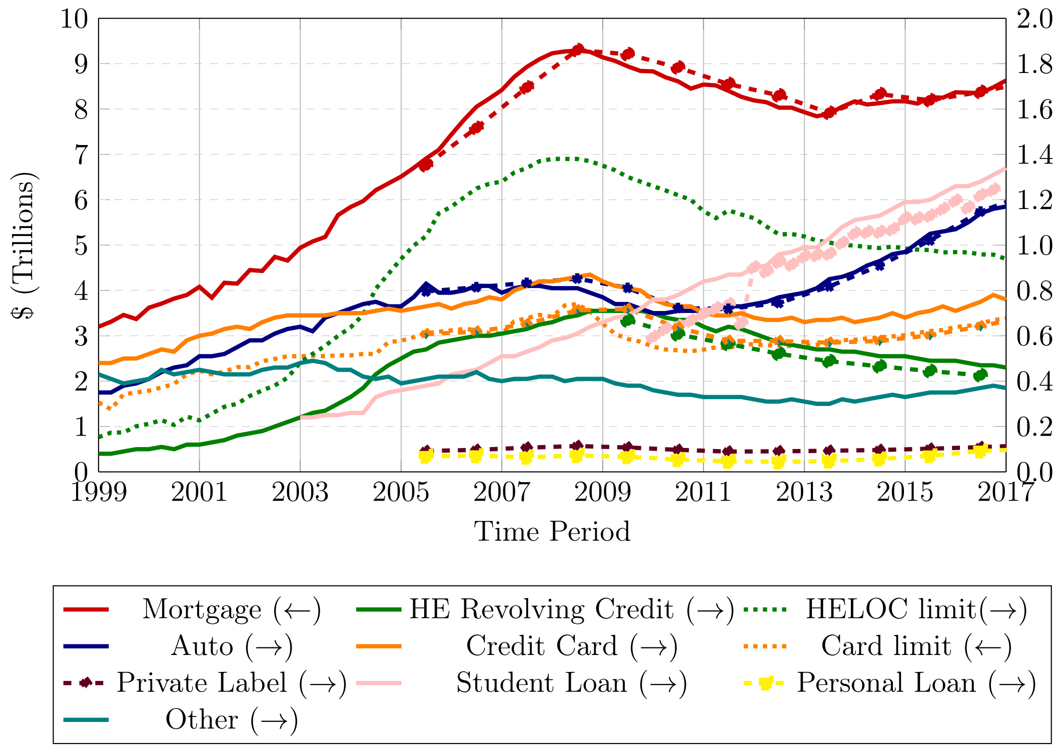


Figure 1.2: Outstanding balance (solid lines), limit or revolving accounts (dotted lines) by quarter and type of credit as reported by the Federal Reserve. The symbols represent the data from the TransUnion Industry Insights Report. The arrows indicate which scale to use.

of accounts or the number of consumers that hold a certain type of credit overall and by unit of time is helpful to study in this regard. The number of accounts is given in figure 1.3 and the number of consumers with a specific type of credit is given in table 1.3. It is also important to be clear on the meaning of the term *Consumer Credit*.

1.2.1 Basics

What is Consumer Credit?

Consumer credit is granted when *money, goods or services are provided to an individual in the absence of immediate payment*. Common forms include credit cards, store cards, motor vehicle finance, personal loans (installment loans), lines of credit, retail loans and mortgages. In a strict sense, debts obtained to purchase margin on investment accounts or real estate are not consumer credit but investments. However, most statistics include mortgages as part of the consumer credit market. The counterparts of consumer credit are e.g commercial credit for businesses or credit for governments while credit granted to non-profit organizations is sometimes included in consumer credit.

How many consumers are there?

For practical reasons we shall limit this consideration to United States consumers. A quick look at the US census bureau gives an idea of how many consumers there are.

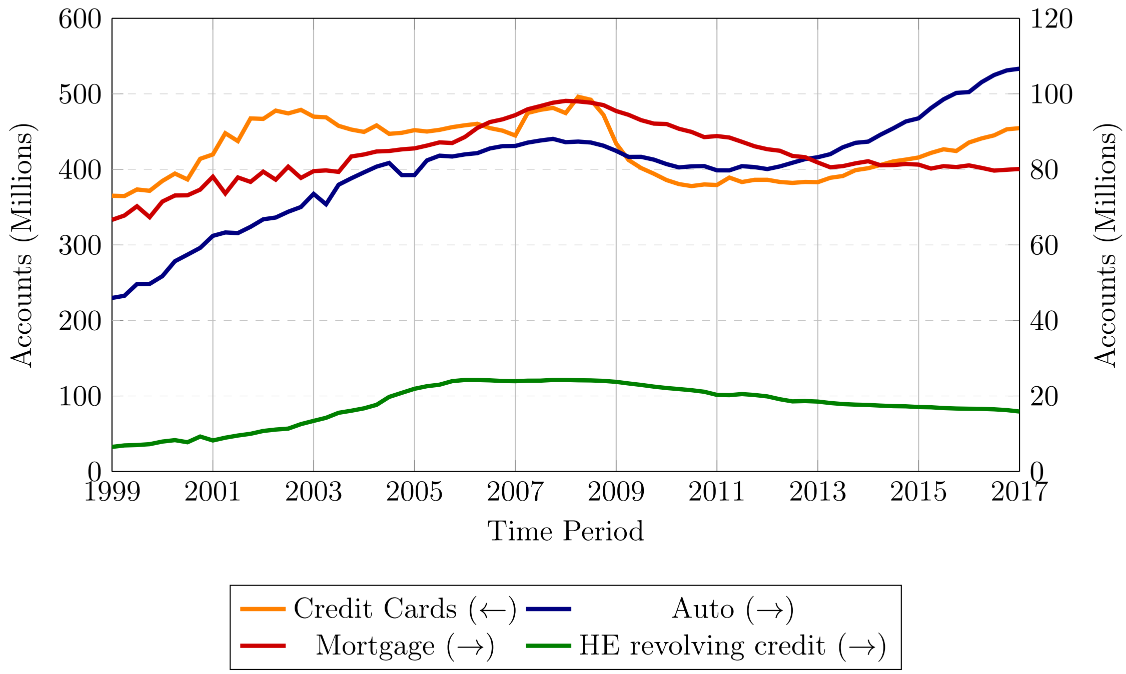


Figure 1.3: Accounts with a balance by quarter and type of credit as reported by the Federal Reserve.

As of July 2016, the US has a

- population of ≈ 323.1 million.
- Roughly 250 million are more than 18 years of age (77.2%), therefore able to contract and obtain credit.
- They live in ≈117M households with an average household size of 2.64 individuals.
- The median income per household is $53,889, the overall income per capita $28,930.
- A total of 124M individuals are currently employed. The unemployment rate is about 4.7% and labor force participation around 63%.
- The total payroll annually is $6.254 trillion. So market size is about 2x the annual payroll with another half of it open to draw.

The credit bureaus have records on

- about 225M consumers, 25M consumers have no file on record.
- Of those with files, about 200M or 89% are scorable by conventional credit scores (see chapter 4).
- 195M consumers have revolving credit lines.
- 142M consumers have non-revolving loans.

The numbers are based on Transunion data. In order to assess the market size by $ volume all we need is statistical data on credit products, an idea on customer demand - in other words which product at what stage of life is needed - and relate that to the population figures.

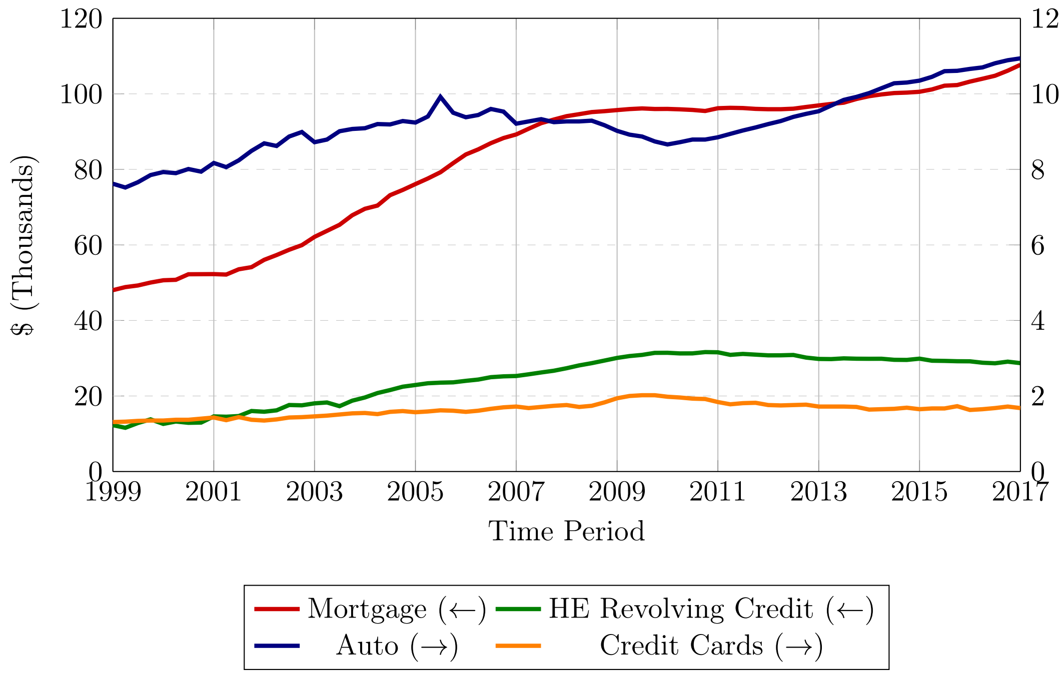


Figure 1.4: Average remaining balance by quarter and type of credit, derived from the data used in figures 1.2 and 1.3 where applicable.

1.2.2 Consumer Credit Market

The consumer credit market is closely monitored by several federal agencies. Both, the Federal Reserve (quarterly status of credit) and the Consumer Financial Protection Bureau (monthly issuance), publish standardized reporting obtained from a representative sample of credit bureau reports. TransUnion provides a bigger overview using the entire data set available (no sampling) and publishes it as the Industry Insights Report every quarter. However, it lacks the standardized reporting over a long period of time of the government documents. Using those sources, we can establish some general perception

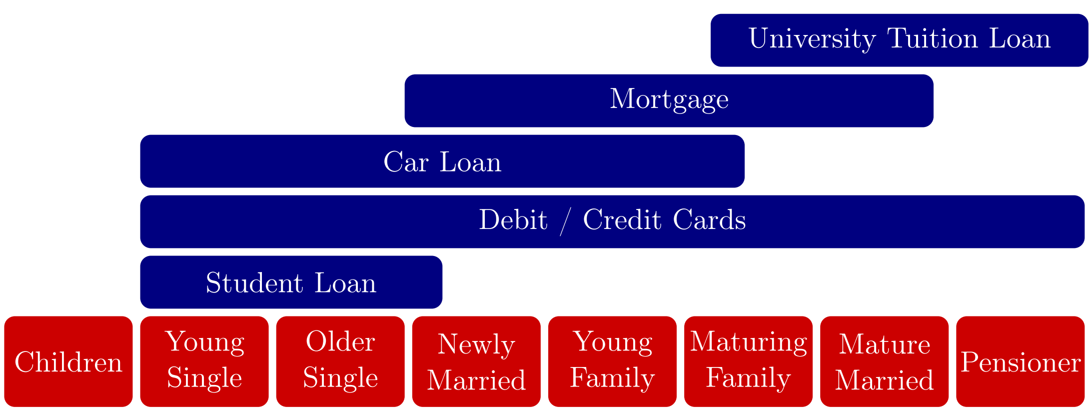


Figure 1.5: Consumer credit needs / opportunities by life stage.

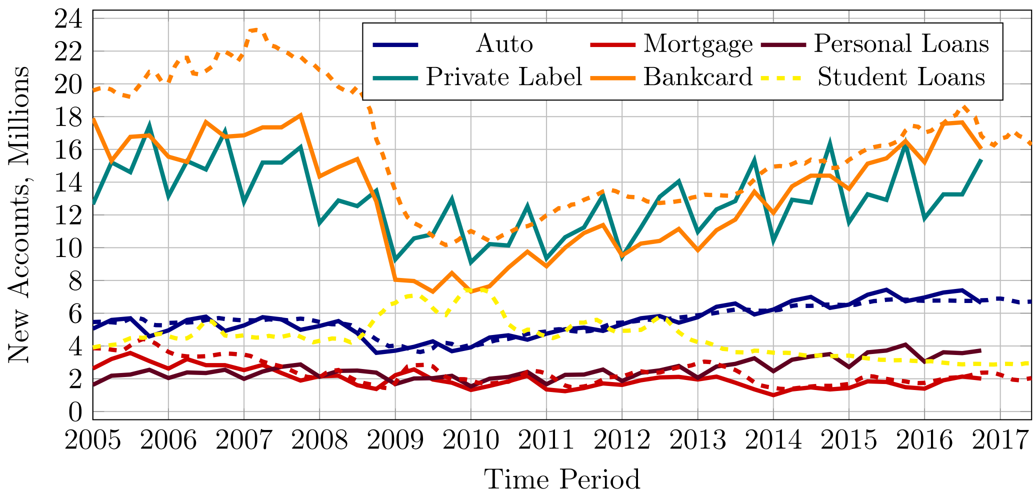


Figure 1.6: New accounts by quarter and type. Solid lines are form TransUnion Industry insights Report, dashed lines from Consumer Financial Protection Bureau.

of market size[6].

Table 1.3 summarizes the state of affairs in Q1 2017. Student loans are not covered in the Industry insights report or other publications due to the fact that obtaining a student loan is not based on any credit data but educational prospects. Only overview numbers are readily available even though it is currently the second largest segment of consumer credit by $ volume and also the one with the highest delinquency rates. Overall estimates are roughly $1.45T for about 44M consumers. About 6.1% of those consumers have delinquent balances which account for up to 11% of balances of student loans that are severely delinquent (**add source**).

The overall balances total around $12T about 2/3s of which are mortgage balances. Per quarter, roughly 53M new credit accounts are opened. Looking at the average new account, the quarterly size of the consumer credit market is in the order of $1T. Personal loans are the smallest, yet fastest growing segment.

An illustration of new account creation over time is given in figure 1.6, showing that the number of account originations holds relatively steady in past few years.

In terms of $ volume issued per quarter,

- Mortgages make up about $300-600B, with high seasonality effects
- Car loans $135B,
- Credit and private label cards $120-140B,
- HELOCs $30-35B,
- Student loans $40B,

[6] The Federal Reserve data provides a sense of the development of credit over time, while on the here and now, the most accurate data point comes from the TransUnion Industry Insights Report. We will use the appropriate data on a case by case basis and if possible, compare them and point to differences. In case of numbers, the TransUnion report has fewer statistical uncertainties and should be regarded as the better source of truth. Both kinds of reports are based on single bureau data and are consequently blind to credit that is not reported to the respective agency.

[7] Product type abbreviates Rev for Revolving credit and Inst for Installment credit.

[8] Avg. means Average, Bal for balance, Cons for consumers

Table 1.3: Key parameters for consumer credit market based on Industry Insights Reports of Transunion Q2 2017. This puts the overall outstanding balances at $12T.

Product	Mortgage	HELOC	Auto	Bankcard	Priv. Label	Personal	Student
Product Type[7]	Inst	Rev	Inst	Rev	Rev	Inst	Inst
Secured	yes	yes	yes	no	no	no	no
Current Status							
Cons. w. bal. (M)[8]	67.9	10.8	81.7	139.8	68.2	16.1	
Cons. w. act. t. (M)	67.9	12.0	81.7	172.6	126.2	16.1	(44)
Accounts (M)	53.0	9.0	77.4	409.6	327.7	17.3	
Accts/Cons.	0.8	0.8	0.9	2.4	2.6	1.1	
Total Balance (B$)	8,726	413.3	1,145	713.6	119.5	106.7	(1,450)
Avg. Balance/ Cons. w. bal. ($)	164,769	58,248	14,802	5,422	1,752	6,161	(32,950)
Avg. Bal./ acct ($)	164,641	45,922	14,793	1,742	364	6,167	
Total Crdt lines (B$)		757.5		3,320	851		
Avg. line/Cons. ($)		98,105		22,745	6,745		
Avg. line ($)		84,167		8,105	2,597		
Quarterly Originations Q1 2017							
Orig. count (M)	1.5	0.3	6.7	15	11.1	2.8	(4.0)
/ US consumer (%)	0.6	0.1	2.7	6.0	4.4	1.1	
/ US households (%)	1.3	0.2	5.7	12.8	9.5	2.4	
Avg. new credit ($)	219,743	112,715	20,415	5,817	2,722	6,430	
Orig. volume (B$)	329.6	30.4	136.8	87.3	30.2	17.9	
of balance/limit (%)	3.78	4.01	11.95	2.63	3.55	16.75	
Serious delinquencies (90^{+}DPD for cards, 60^{+}DPD all others)							
Consumers (%)	1.92	0.62	1.23	1.46	1.25	3.02	6.14
Accounts (%)	1.97	0.72	1.07	0.83	0.63	2.99	
Balances (%)	1.97	1.47	0.95	1.36	2.16	1.79	

Personal loans around $20B.

Personal loans are less than 1% of current market, but more than 2% of new issuance and this segment is growing fast.

The overall net account growth is in the order of 1.8M per quarter as derived from the Federal Reserve Data in figure 1.3, driven mostly by credit cards. This also implies that with 53M accounts originated by quarter and a net addition of 1.8M, that 51.2M accounts must have been paid off to a zero $ balance, charged off or closed in the same time period.

American dream built on installment credit

1.2.3 History

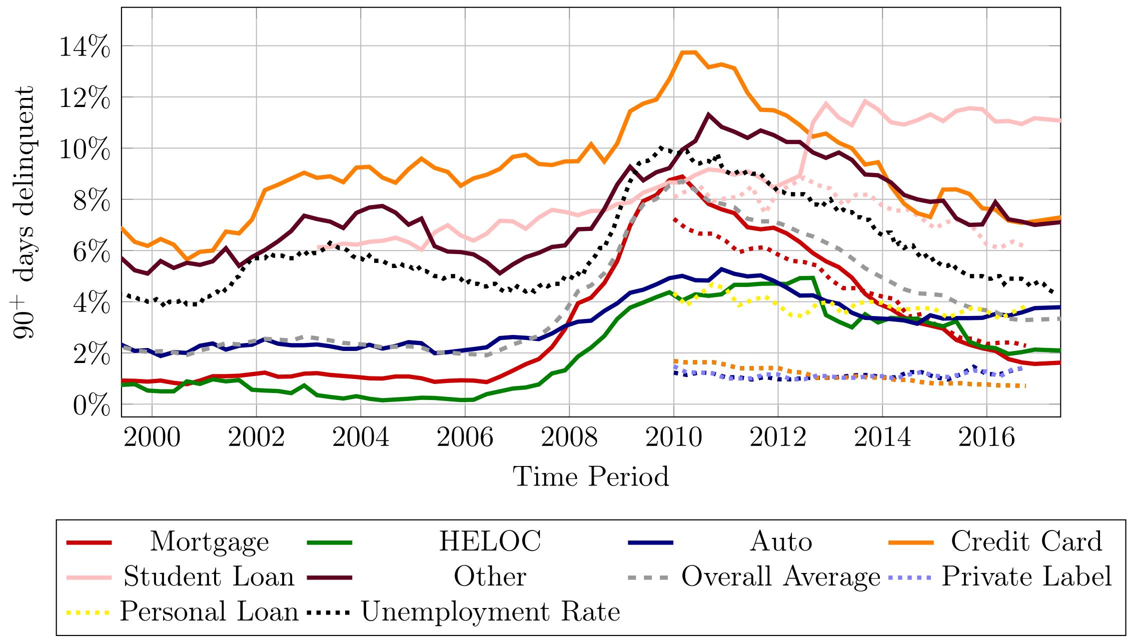


Figure 1.7: Percent delinquent balance by quarter and type of credit as reported by the Federal Reserve (solid lines). Dotted lines indicate delinquent consumers in category as given in Transunion's Industry Insight Report in corresponding color.

1.2.4 Performance & Repayment Behavior

In every portfolio some consumers fall behind on their regular required payments for the credit obtained and thus become *delinquent*. After 90 days, sometimes 60 days, a delinquency is considered severe. A charge-off usually occurs after 120 days or 180 in the credit card space. Table 1.3 provides numbers for average delinquencies in Q2 of 2017 by product type. The expected severe delinquencies / charge-offs / bankruptcies are key to the design of any financial / credit product (see section 1.3). Low delinquencies indicate responsible market growth, as it would be all too easy to give a lot of money out to grow, but take unfathomable risks by doing it indiscriminately.

Delinquent balances are part of standardized reporting by the Federal Reserve. While delinquent balances or account numbers are determined relatively straightforward, several ways exist to analyze relative delinquencies, count based incidence or money based measures:

- Delinquent balance as percentage of outstanding balances (e.g. given by the Federal Reserve of New York's Quarterly Report on Household Debt and Credit)
- Delinquent balance as percentage of original amount lent (this would be interesting to financial models and risk)
- Accounts delinquent as percentage of accounts originated
- Consumer with delinquent accounts as percentage of consumers with accounts of a specific type (e.g. TransUnion Industry Insights Report)

Figure 1.7 shows the average delinquent balance 90^{+}DPD as reported by the Federal Reserve of New York's Household Debt and Credit report as well as the percentage of consumers with delinquent accounts 60^{+}DPD (credit cards 90^{+}DPD) of consumers

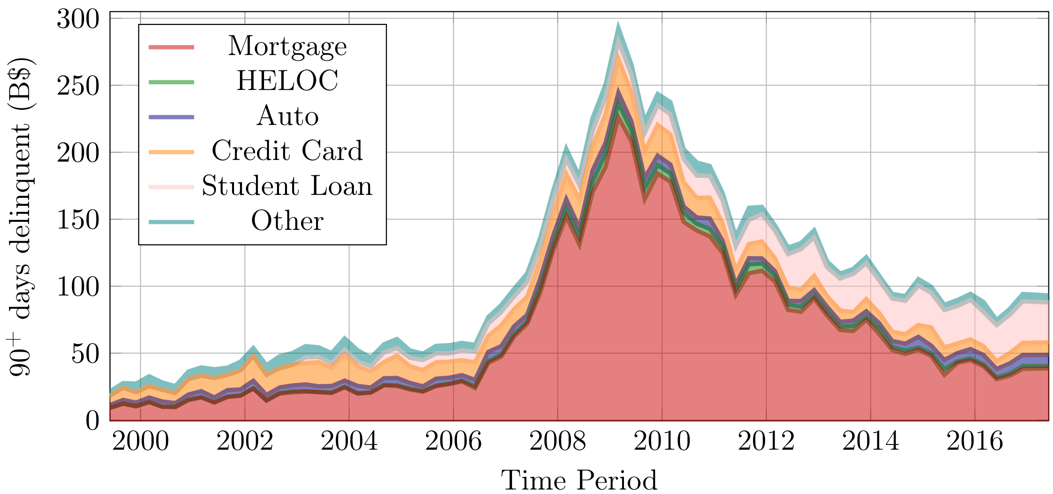


Figure 1.8: Delinquent balance by Quarter and type of credit as reported by the federal reserve.

holding the account type as reported by the Industry Insights Report by TransUnion . There are a few things to note here,

- Both groupings of curves show similar trends.
- The delinquent balance ratio does refer to balances at the observation point in time, not issued volume. It ignores prepayments and principal pay-down that have occurred post issuance but prior to the observation. The ratio will therefore overstate what delinquencies on new credit could be expected.
- The balance based delinquency measure is larger than the consumer count based one in all cases despite the fewer days past due in this statistic, largely thanks to two things: 1) a delinquent consumer with several accounts may have become so on more than one account, 2) Delinquencies tend to occur more frequently on higher balances.

Figure 1.8 shows delinquent balances as per the Federal Reserve of New York. This illustrates once more the dominant position mortgage debt has in the consumer credit space.

According to Transunion 2.23 million consumers declared bankruptcy in 2016. This down from 6M in 2010. Around 70% of consumers that declare bankruptcy do so using chapter 7 (liquidation), the rest chapter 13 (reorganization) or the type of bankruptcy filing is not known. The same source puts the number of charged off trades by year to 20M on a $46B charged off balance (2010 : 26.5M trades for $93B). Nearly half of the overall charge offs occur in Bankcards by both count and balance.

At present, bankruptcies by count are roughly 1.0% or new accounts generated and the charge-off balance holds at around 1.2% of new balances.

1.3 Underwriting 101

From Wikipedia:

Underwriting *services are provided by financial institutions whereby they* ***guarantee payment*** *in case of damage or financial loss and* ***accept the financial risk*** *for liability arising from such guarantee. An underwriting arrangement may be created in a number of situations including insurance, issue of securities in primary markets, and in bank lending, among others.*

The name derives from the Lloyd's of London insurance market. Financial bankers, who would accept some of the risk on a given venture (historically a sea voyage with associated risks of shipwreck) in exchange for a premium, would literally write their names under the risk information that was written on a Lloyd's slip created for this purpose.

In short, underwriting assesses the acceptable risk to required premium ratio.

1.3.1 Classic Underwriting

In classic (old school) underwriting the five Cs are commonplace. A good summary can be found in this PWC report for both business and consumer underwriting.

Character Credit Score (mostly FICO), Credit History (reports), Time at Address, Time at Job, Time in Business, Industry, Business Cash Flows, Educational Background, Experience of Employees, References

Capacity Debt-to-Income-Ratio, Debt-to-Equity-Ratio, Past Income, Employment History, Type of Income

Collateral Net worth, Liquid Assets, Equipment, Buildings, and Inventory, Accounts Receivable

Capital Assets Pledged as Collateral, Retained Earnings, Personal Guarantee, Co-signer Guarantee

Conditions Local Economic Trends, Industry Trends, Competition, Loan's Purpose

The white paper also illustrates how technology or social media data in particular can be used to strengthen the underwriting decision confidence.

What are specific points peculiar to Fintech underwriting? Oftentimes, Fintech enabled companies focus on the credit score and history parts of the underwriting decision. This constitutes an important but not the only major part of underwriting. The strength of technology enabled companies is additional, automated verification processes around all other parameters of the five Cs as well.

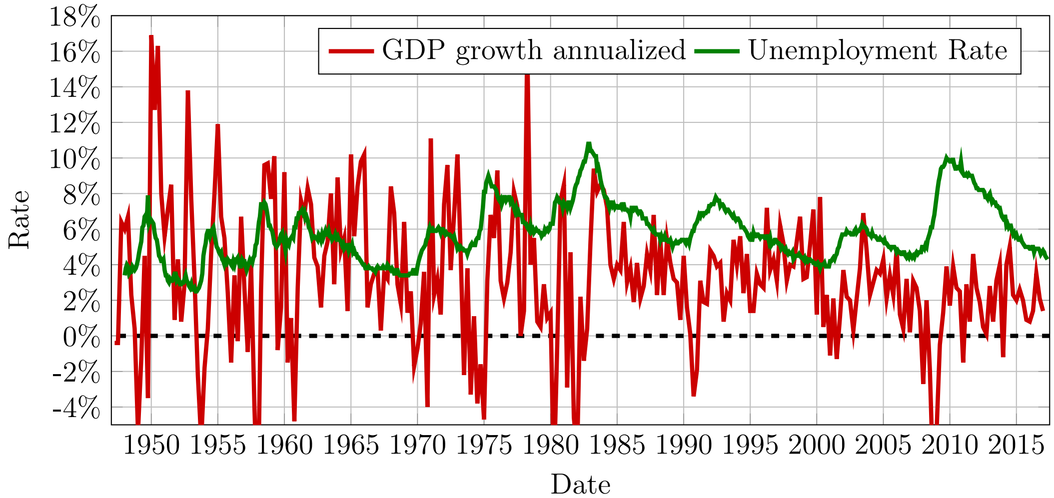


Figure 1.9: GDP growth and unemployment rate. From the Federal Reserve

1.3.2 Fintech Underwriting

This passage is modified from a post by Frank Rotman, FinTech Junkie Blog .

Next Generation / Online Lending has been around for roughly a decade that coincides with fairly benign economic period characterized by sustained economic growth (see figure 1.9), low default (see figure 1.10) and interest rates (see figure 1.11). There is considerable scepticism about how the models will perform in a downturn and how the portfolio's performance can be protected against a sudden and massive deterioration.

It boils down to the question:

- ***How resilient is your borrower base?*** and therefore understanding of
- ***How many things have to go wrong?*** at the customer level.

Credit is all about trust[9]. So lenders will make loans to healthy borrowers that they believe are willing **and** able to pay back the loans over the time period in question. Unfortunately, a borrower's circumstances change over time and that matters. The result is a fraction of borrowers in *every* portfolio that will default on their loans.

A healthy borrower has the following traits

- A relatively stable source of income that supports one's obligations/lifestyle
- Enough savings to weather a temporary disruption to one's income
- Enough savings or free cash flow that can handle the introduction of additional unforeseen expenses
- A willingness to pay one's debtors when the money is available
- The ability to quickly find a new source of income after a disruption
- The ability and willingness to turn collateral into cash to pay one's obligations

As the Cynic's Guide to Fintech so poignantly remarks, *...banks almost never lose money on bad risks. They lose money on good risks, which go bad.*

[9] That is the literal meaning of the word **credit**.

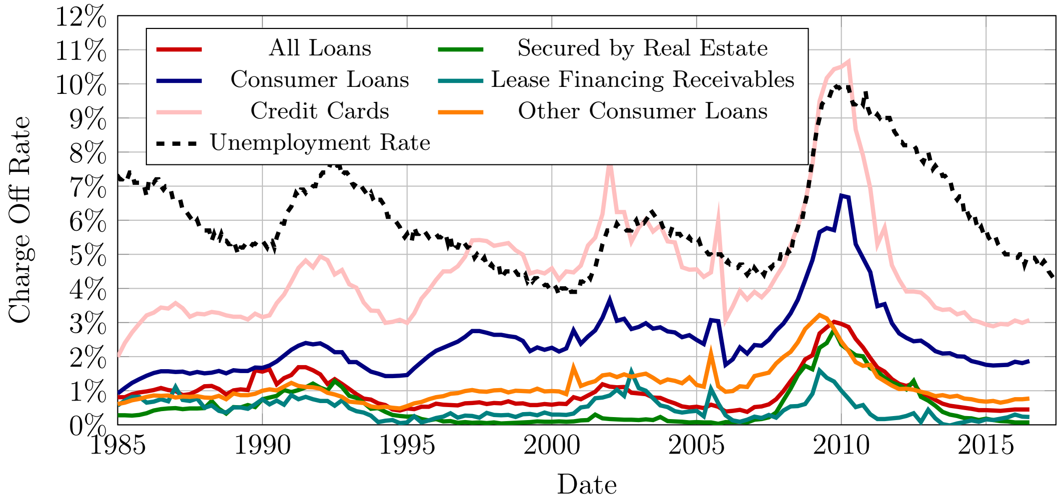


Figure 1.10: Delinquency rates reported on commercial bank loans according to the Federal Reserve, seasonally adjusted. A more detailed look at the recent deliquency history is given in figures 1.7 and 1.8.

A breakdown of a healthy customer can be traced to a fundamental change in their circumstances.

- Temporary reduction of income (job loss, reduced commission, etc)
- Permanent reduction of income (major change in health, retirement, etc)
- Increased cost of living (increased borrowing, new child, etc)
- Unforeseen major expenses (car repair, medical bill, etc)
- Reduction in financial safety net (increased spending, reduction in new savings, etc)
- Reduction in willingness to pay (strategic default, attitudinal change, etc)

Statistically based underwriting models perform better than humans because models are able to predict the natural change in circumstances at the customer and portfolio levels. A model does not classify a customer as *good* or *bad* but rather that they have a certain probability of paying back a loan.

But both human and statistical based underwriting models/policies suffer from the same phenomenon - tomorrow is not guaranteed to look like today. And while models are able to project an ambient deterioration in a portfolio's performance, they are not fundamentally able to project what will happen in a future they have never seen before.

The natural reaction from investors and entrepreneurs who haven't managed loan portfolios through cycles is to be terrified of what's to come. Investors want to naturally stop investing in companies that originate loans. Less experienced entrepreneurs don't know how to build resilient underwriting models and convince investors that all is *OK*.

Therefore, it is of paramount importance to make sure your models give significant weight to the major drivers of risk (ability to pay, willingness to pay, stability of income, etc). Just exposing a model to hundreds of potential variables is not good enough. All models must appropriately weigh each and every potential driver. If one is missing due to modeling techniques, it's use must be forced into the models or policies in order for the

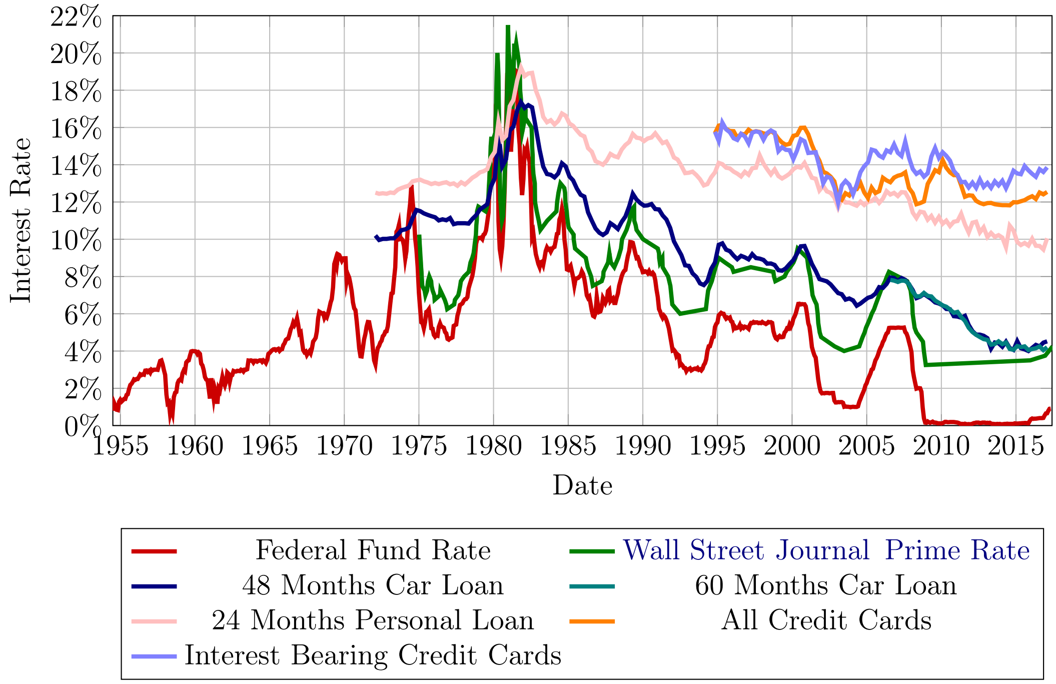


Figure 1.11: Interest Rates history for all reported commercial bank loans according to the Federal Reserve, Prime rate from the Wall Street Journal, historic data via HSH Associates. Data is not seasonally adjusted. It is worth noting and driving the development of new players that the spread between the federal funds / prime rate and e.g. credit card rates has been increasing in the last decade.

.

portfolio to stay resilient against things that go wrong within the customer base while not exceeding expectations on defaults.

For example, it's critical to avoid lending to customers who are on edge from a capacity standpoint because minor changes in their circumstances will push them from solvency to insolvency. DTI[10] *might not show up in the underwriting models because our economy has been great for the past handful of years, but I can definitively say from experience that DTI won't matter until it's the only thing that matters. Models don't understand this.*

Make side note with bankruptcy stats akin to the Warren post, something like 3 reasons, divorce, job-loss, medical, with numbers, find official reference, additionally, some comments on strategic defaults of underwater consumers, think there is a fico reference

[10] Debt-to-Income Ratio

1.4 Funding for lenders 101

A lending business is a cash hungry enterprise and solid funding is a necessity. There are three fundamentally different approaches to funding:

- the deposit model,
- the balance sheet model and
- the marketplace model

Deposits are a highly regulated source of funding available to chartered banks only. It is not a feasible option for a startup. In the balance sheet model a lender employs its own money (also know as *skin in the game*) and is therefore less likely to originate borderline borrowers. The marketplace model on the other hand can be used to build a lightweight and scalable technology company. A hybrid model combines both and helps to maximize the diversity of funding sources to better weather the next liquidity crunch in any funding source. **AVANT** employs a hybrid funding model.

1.4.1 The Balance Sheet Model

In the balance sheet model the lender uses its own balance to fund loans. This is a business model that has been tested through time and was prevalent before technology began to take over. The goal here is to lend at a high rate and borrow at a low rate to capture the maximum spread. The spread is driving the revenue. Technology here drives efficiencies in the lending operations, underwriting processes using machine learning. Traditionally, specialty finance companies were viewed as low-margin, low-growth, low-multiple business. When combined with technology, it yields double to triple the Enterprise Value-Revenue multiples than non technology enabled competitors (see the hybrid model blog post).

What funding sources do we find?

Venture Capital

Most lending startups first employ venture capital to fund loans. VCs typically take a high risk for a high return. The likely target IRR is in the order of 30% which will be hard to achieve with unlevered loans that yield 12%. However, they are willing to provide the first equity which will eventually be levered once transitional capital providers get on board. At start, the capital will be used to fund loans directly and will be reinvested after a short period of time and the onboarding of new capital providers in more traditional ways (hire new staff, invest in better technology, acquire new borrowers, etc.).

Transitional Capital

Transitional Capital is the first pure lending capital provider. They take early stage lending risk in exchange for unusually high returns, typically high teens, low twenties. They evaluate the management team, credit and underwriting models and the early performance of loans. As they are high priced providers they will be transitioned out to lower cost providers once the track records firm up (hence the name). It is not uncommon to cycle through several providers of transitional capital to progressively bring the cost of capital down.

Warehouse Lines

A Warehouse line is a huge line of credit that the originator can draw from typically provided by banks and large funds. They will require an established business model, proven track record and consistent underwriting. Basically, no surprises. As a result they are willing to offer their capital at a lower rate of return than Transitional Capital providers. Over time, with longer track record, warehouse lenders usually offer more attractive advance rates and lower cost of capital.

Securitization

The Securitization market provides two major benefits when it is working. First, a low cost of capital from insurance funds, pension funds and other large allocators, giving the lender a healthy spread between rates charged and rates paid for capital. Second, it allows the originator to quickly recycle their capital allowing it to originate more loans with a limited amount of capital. This is key on the way to $Bs a year originations.

1.4.2 The Marketplace Lending Model

The marketplace lending model is way of creating a lightweight online financial market that connects lenders with borrowers. The lender usually takes care of pricing using technology. The revenue stems from transactional fees from the borrower and a servicing fee from the investor. Marketplace lenders generate much lower revenue than balance sheet lenders, yet at a substantially higher margin. The marketplace lending model allows a new category of investors to participate in the funding.

Whole Loans

Whole loans are offered to accredited investors by the capital markets team. Typically they are offered to P2P Lending funds, high net worth individuals and family offices. The marketplace model is usually added after a minimum scale is achieved using a balance sheet. It is common for marketplace lenders to have their own whole loan index fund to purchase loans off their own platforms.

Levered Funds

In the balance sheet model your capital markets team negotiates for their own warehouse line. In the marketplace model your capital markets team negotiates with the same providers to make their capital available to fund managers that seek leverage for their funds. The capital providers must complete similar due diligence on the marketplace as they would on the balance sheet originators but in addition, they must also evaluate the fund manager. Many of the leading P2P funds managers offer both unlevered and a levered funds including Colchis, Eaglewood, Incline, Echelon, and HCG to name a few.

Sponsors

Some fund managers act as securitization sponsors. The fund manager secures a warehouse line and applies it to a credit facility. Once the facility is filled with loans, the sponsor securitizes the facility into two or three tranches and sells the newly formed securities to insurance companies, pension funds, and large credit funds. Recently, the marketplaces have decided to take a more active role in this process. The marketplaces are working toward a new model where the originator files the shelf and acts as the sponsor and then coordinates with several fund managers that offer their credit facilities with standardized terms through the Marketplace Lending shelf.

Fractional

The fractional loan market is designed for smaller investors that cannot achieve proper diversification through the whole loan market. This market includes smaller fund managers, high net worth individuals, and family offices as well as wealth managers and self-directed investors. The fractional market is designed for the retail investor universe, which is highly fragmented and difficult to aggregate.

The behavior of this group of investors is much different than the institutional market because of their long term time horizon and different risk appetite. In many cases, retail capital can serve as counterbalance to institutional capital to stabilize investor inflows and outflows. Online lending is an ideal product for retirement accounts since it offers stable returns, high yields, low correlation to the equity and bond markets, and short duration. Up to 40% of all loan originations could eventually come from retirement accounts through the wealth management channel once the infrastructure is in place. (see the Hybrid Funding blog post)

1.4.3 The Hybrid Model

We are seeing a convergence of the balance sheet and marketplace models. In the balance sheet case the platform is focused on profitability/spread per loan; in the marketplace model the platform is focused on volume but profitability per loan is small. Neither approach seems to be perfect. The balance sheet approach is capital intensive, the marketplace model requires massive volume. The hybrid model helps to solve these problems.

Many of the leading originators are moving towards the hybrid model. **AVANT** has been very vocal about its hybrid strategy and has been a leading example for the industry. As platforms are able to diversify their capital sources the hybrid model is likely to prove to be the most sustainable business model over the long term.

Finally, it is worth noting that the ultimate source of capital are the banks themselves. If you can sell directly to a bank, you can effectively operate with 2.5%-4.0% cost of funds without all the overhead of a securitization program or the need to hold the residual strip. Companies like GreenSky do this well in the consumer category. The auto and mortgage sectors are other categories with originators that actively sell directly to banks. If you have a product that allows you to write loans to a bank's credit box, and you have the scale and track record, then selling loans directly to a bank will get you a very low cost of capital.

possible graphs:

- **Funding volume**
- **Securitization volume**
- **table of key providers in each funding**
- **table of market with date founded, market (US/ other) , finance model, target product**

1.5 Fintech 101

- telegraphs
- Automatic clearing house
- Automated Telling Machines
- Online systems
- Expand here with years and quotes

1.5.1 Lending as a service

1.6 Further Reading and Data

Further study of the respective blogs / data is strongly recommended.

Reference & Literature:

- The Hourglass Effect: A Decade of Displacement, Frank Rotman, Peter Renton, April 10, 2015
- Is it time for consumer lending to go social?, PWC, February 2015
- Belts and Suspenders, October 29th 2016, Frank Rotman, FinTech Junkie Blog
- A Cynic's Guide to Fintech, Dan Davies, Bull Market, Medium, April 2015
- Quarterly Report on Household Debt and Credit, February 2017, Federal Reserve Bank of New York, This is a 5% sample of the population with credit files provided by Equifax.
- Consumer Credit Trends, Consumer Financial Protection Bureau, this too is a panel based on credit files
- Industry Insights Report Q1 2017, February 2017, TransUnion
- Effective Federal Funds Rate, Percent, Monthly, Not Seasonally Adjusted, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis, April 2017
- Delinquency and Charge Off Rate, Percent, Monthly, Seasonally Adjusted, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis, April 2017
- HSH Associates, Prime rate archives, May 2017
- Jason Jones, The Pure Marketplace Lending Model is Dead, the Hybrid Takes its Place, September 26th, 2016

Chapter 2

Installment Loans

An installment loan or *loan* for short, refers to a form of fixed term credit. Sometimes it is called an Installment Plan or Hire Purchase. A loan is a debt provided by the lender to the borrower evidenced by a promissory note or contract that specifies among other things the principal amount of money borrowed, the interest rate charged by the lender and the date(s) of repayment.

A loan entails the reallocation of the subject asset(s) for a period of time, between the lender and the borrower. In a loan, the borrower initially receives or borrows an amount of money, called the principal, from the lender, and is obligated to pay back or repay an equal amount of money to the lender at a later time.

Key features / Characteristics of a loan are:

- A fixed principal amount,
- A fixed length of term,
- A pre-set number of payments with specific, typically equal amounts to be paid each installment. This is referred to as a(n installment) schedule.

The lender will provide the funds at the beginning of the loan (draw down) and the consumer will then pay back smaller amounts. The amount borrowed as well as the amount of payments are not subject to change in the course of the life of the loan.

While **AVANT** currently offers only unsecured installment loans, the mechanics of loans discussed in this chapter apply similarly to all installment loan types, just differ in the pricing and collaterization of the loan. An overview of installment loan types is given in figure 2.1.

In Figure 1.2, it is shown that the bulk of outstanding consumer credit is in the form of installment loans with mortgages making up the lion's share of ≈\$9 trillion (≈68%). Other large loan balances are on car loans (\$1.2 trillion, ≈9%) and student loans (\$1.4 trillion, ≈ 10%), all of which work with the same underlying mechanics as discussed in this chapter. The rest of the balances are 6% credit card balances (≈\$760B), 4% home equity line balances (≈\$460B) and 3% of other credit balance (≈\$370B). Personal Loans as offered by **AVANT** are a contributor to the overall balances that is steadily growing but on a comparatively small volume.

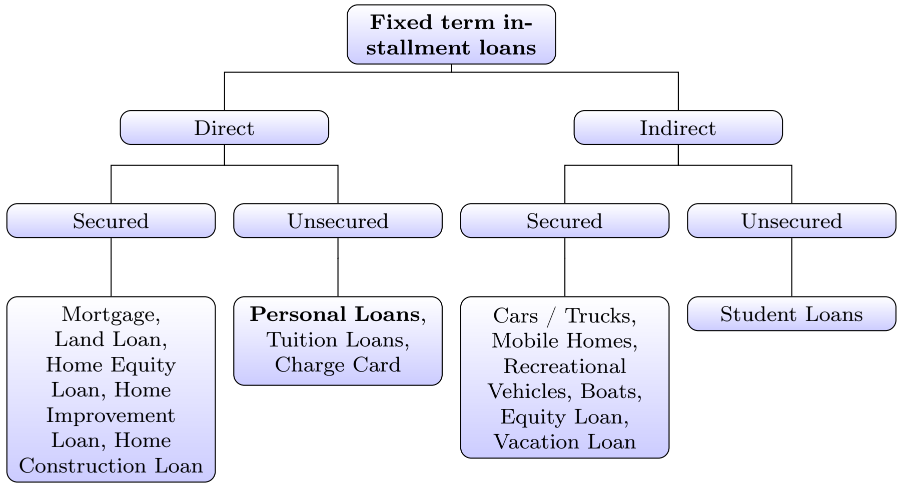


Figure 2.1: Types and hierarchy of fixed term installment loans.

AVANT issues unsecured, fully amortized[1], simple-interest[2], fixed-rate[3] loans of equal payment amounts at each installment. As long as the customer stays on the happy path[4] **AVANT** does not apply fees or charges other than interest after origination[5].

To determine the details of a loan we and the customer are interested in the following:

- What is the **amount** of payment per month?
- **How much** of that payment is **interest** vs. **principal**?
- How much **total interest** is to be paid and how does it relate to **APR**?

As **AVANT** we would like to know a few more numbers to develop a profitable business

- **How much** money do we make?
- **When** do we start making money (after which payment) for each loan in the best of cases (happy path)?
- **How much** money do we make in each loan class if he doesn't?
- **How many** loans of the same type need to be paid off fully in order to maintain a profitable product given a delinquency on any given payment?
- **How many** loans of the same type need to be paid off fully in order to provide a specific return given delinquency, prepayment and over payment behavior as well as cost of funding, underwriting, issuing and acquisition of customer?

[1] Amortization refers to the rate at which the initial amount borrowed (principal) is reduced, fully amortized loans have you pay the exact same amount each month, paying off both interest and principal. Other examples are interest only loan payments with cheaper initial payments during the so called interest only period. After that you pay interest and principal and pay a lot more than on fully amortized loans

[2] Interest is only calculated on outstanding principal, not on principal and outstanding interest as in compound interest.

[3] Fixed-rate refers to unchanging interest rate as opposed to adjustable interest rate. During the life of a loan the monthly payment for fixed-rate loans will not change as long as you pay on time. Adjustable-rate loans adjust the interest rate to current market standards at specific adjustment periods, which means you can end up owing more or less money after changes of interest rate.

[4] Customer pays all his rates in full and on time.

[5] While **AVANT** originally did not have origination fees, those have been introduced in 2016.

2.1 History and Market Overview

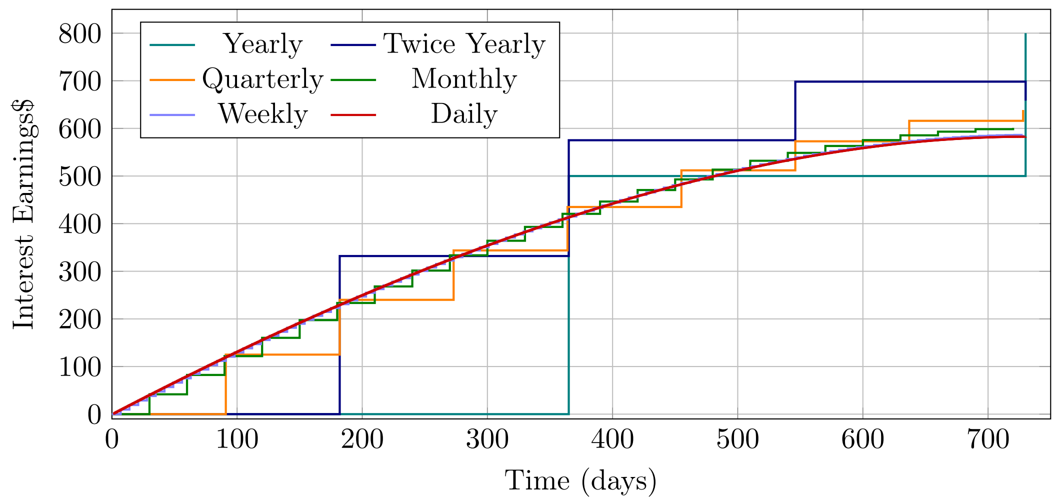


Figure 2.2: Effect of the payment interval on the time accrual of interest. Given for A=50%, 2 years and P=$1,000.

2.2 Consumer Considerations

In this section we shall assume that we have a customer who

- is **rational**,
- aims for an optimum (**minimum**) of loan **cost**,
- does **not** become **delinquent** on any payment (to be discussed in section 2.3) and
- has the financial & technical capacity to **make all payments**, over and / or pre-pay on time or prior to his due date.

So, given an interest rate, term and maximum borrow amount offer, how should the customer determine what makes financial sense to him? We will derive the loan amounts, interest rates and terms that can be offered to a set group of customers in section 2.3.

Table 2.1: Effect of the number of days d in the payment / interest calculation interval on the payment amount α, the overall amount paid β and the cost of credit γ. $\kappa = \gamma/P$ given for comparison. Numbers are given for a loan with A=50%, 2 years term and P=$1,000.00. See fig. 2.2

Payment Interval	d/#	m	α ($)	β ($)	γ ($)	κ
Daily	1	365	2.17	1,582.61	582.61	0.58
Weekly	7	52	15.25	1,586.40	586.40	0.59
Biweekly	14	26	30.59	1,590.81	590.81	0.59
Monthly	30	12	66.71	1,601.06	601.06	0.60
Quarterly	91	4	204.83	1,638.66	638.66	0.64
Twice a year	182	2	423.44	1,693.77	693.77	0.69
Yearly	365	1	900.00	1,800.00	800.00	0.80

2.2.1 Payments & Overall Cost of Credit

Let us get to it. Let P be the principal amount, A the nominal interest rate[6], n the number of payments (length of term) and m the number of payments in a given year, typically $m = 12$. All formulas are based on the effective interest rate I which in case of monthly computation gets calculated as

$$I = A/m \tag{2.1}$$

Here it is worth noting that we would like this number to be used as a decimal, not a percentage[7].

Side Note:

At **AVANT** , we compute interest daily, but book monthly installments. For that purpose we would have to divide A by the number of days in a year[a] to find the effective daily interest rate I'. A higher payment frequency works in favor of the customer and diminishes interest earnings, therefore computing daily interest is mandated by law. The differences between monthly and daily are not huge but well discernible. An example for the effects is laid out in table 2.1 and illustrated in figure 2.2.

[a] This number is either 366 days in leap years and 365 in non-leap years, 365.25, 365 or 360 days in any given year depending on applicable legislation by state and geography.

For each payment k of equal amount α we can determine the remaining principal P_k

[6] Which in the case of **AVANT** in the US is practically the same as the annualized percentage rate (APR) as long as no origination fee is charged. Disclosure of this term is a legal requirement of customer protection and its determination varies by country. Resource: FDIC APR Simple Interest, FDIC APR Compound Interest

[7] I also wants to be smaller than one, otherwise your payment amounts exceed the original principal borrowed. That means A wants to be smaller than $100m$ if given in percent, if there are m=12 installments per year, that means $A \ll 1200\%$

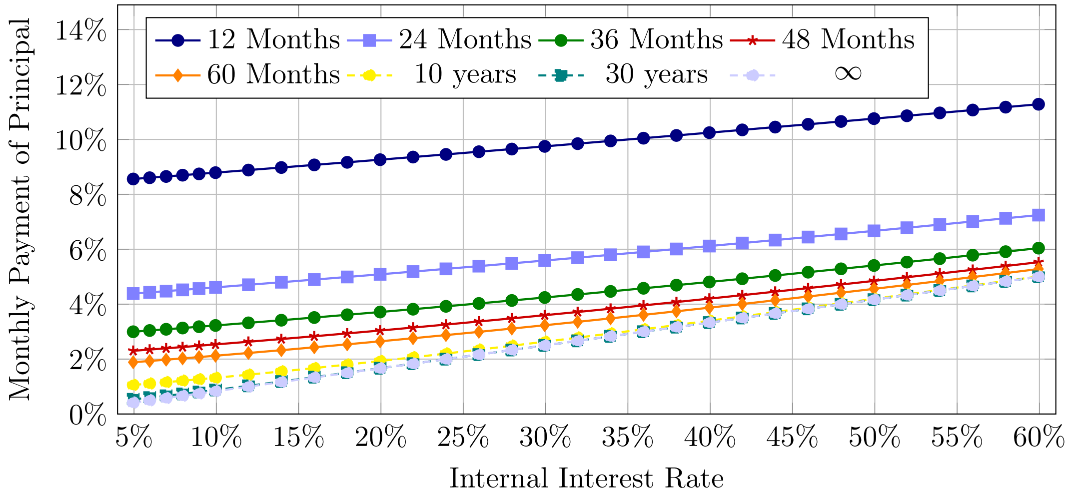


Figure 2.3: Monthly payment as a percentage of principal borrowed for internal interest rates from 5% through 60% for various terms. Apparently the decrease in monthly payment is largest from 12 to 24 months terms and is less appreciable afterwards.

to be:

$$\begin{aligned}
P_1 &= P(1+I) - \alpha \quad , \\
P_2 &= P_1(1+I) - \alpha = [P(1+I) - \alpha]\,(1+I) - \alpha \quad , \\
&= P(1+I)^2 - \alpha\,[(1+I) + 1] \\
P_3 &= P_2(1+I) - \alpha \\
&= \{[P(1+I) - \alpha]\,(1+I) - \alpha\}(1+I) - \alpha \quad , \\
&= P(1+I)^3 - \alpha\left[(1+I)^2 + (1+I) + 1\right] \\
&\ \ldots \\
P_k &= P(1+I)^k - \alpha\left[(1+I)^{k-1} + \cdots + (1+I)^2 + (1+I) + 1\right] \quad . \qquad (2.2)
\end{aligned}$$

We then can transform equation 2.2 using the geometric series computation

$$1 + x + x^2 + \cdots + x^n = (x^{n+1} - 1)/(x - 1) \qquad (2.3)$$

into

$$P_k = P(1+I)^k - \alpha\frac{(1+I)^k - 1}{(1+I) - 1} \quad . \qquad (2.4)$$

Knowing that we want to have the loan paid off after n payments, in other words $P_n = 0$, we can then compute α from

$$P_n = P(1+I)^n - \alpha\frac{(1+I)^n - 1}{(1+I) - 1} = 0 \qquad (2.5)$$

as

$$\alpha = P\frac{(1+I)^n\,(1+I-1)}{(1+I)^n - 1} = P\frac{I}{1-(1+I)^{-n}} \quad . \qquad (2.6)$$

I find it helpful to use as few polynomial expressions as possible to describe a quantity which is why going forward I will use the right hand side of expression 2.6 for α.

An illustration of the relation α/P is given in figure 2.3.

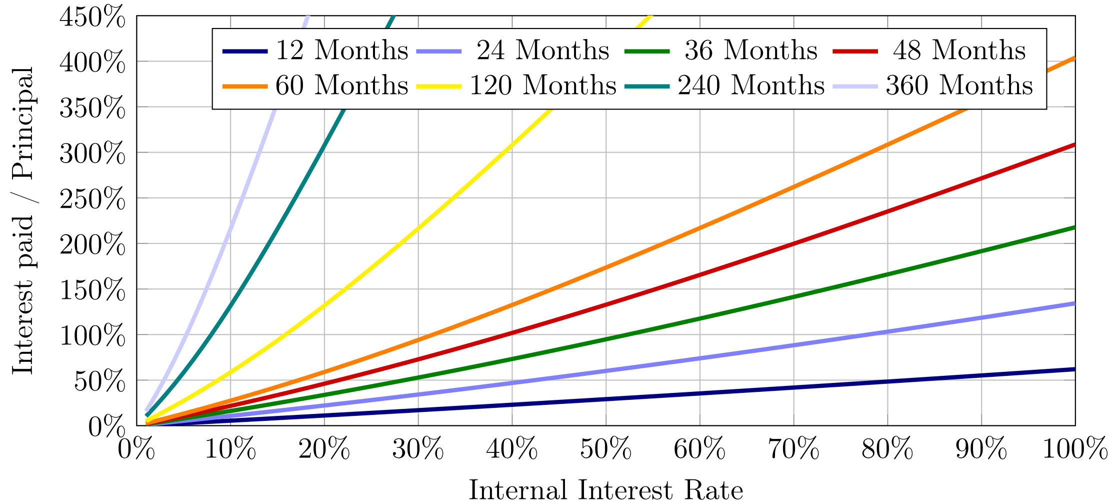


Figure 2.4: Overall Interest paid as a percentage of Principal.

Side Note:

For simplicity we will give formulas here based on monthly computing assuming months of equal length (equal number of days per month). If we instead calculated daily computing, but charge installments monthly, we could substitute I with $I' \times d$ where d is the number of days between installments. As can be easily seen, the monthly payments and overall payment will not change, however, the composition of interest versus principal (see below) may well be experiencing small but appreciable changes. In practice, banks do not operate on weekends and holidays, which can result in a spread for the number of days between payments and the number of days in that month. We will go through the effects at a **get reference** later section.

Limits & Approximations:

Obviously the limit of α for $n \to \infty$ is $I \times P$

$$\lim_{n\to\infty} \xi = \lim_{n\to\infty} \frac{\alpha}{P} = I = \frac{A}{m} \quad , \tag{2.7}$$

which makes a lot of sense.

Using the Taylor Expansion of any infinitely differentiable function f

$$\begin{aligned} f(I) &= f(I_0) + (I - I_0) \times \frac{d}{dI} f + \frac{(I-I_0)^2}{2} \times \frac{d^2}{dI^2} f + \dots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(I_0)}{n!} (I - I_0)^n \quad , \end{aligned} \tag{2.8}$$

on the term $(1 + I)^{-n}$ using a reference to $I_0 = 0$:

$$(1 + I)^{\pm n} = \sum_{n=0}^{\infty} \frac{I^n}{n!} \frac{d^n}{dI^n} (1 + I)^{\pm n} \quad . \tag{2.9}$$

Explicitly, we obtain up to the cubic term

$$(1+I)^{\pm n} = 1 \pm I \times n + I^2 \frac{n(n \mp 1)}{2} \pm I^3 \frac{n(n \mp 1)(n \mp 2)}{6} + \dots \tag{2.10}$$

which for $I \ll 1$, or more precisely $nI/2 = nA/2m \ll 1$ can be used to approximate α in an easier to handle expression.

Up to quadratic term equation 2.10 leads to

$$\alpha \approx \frac{P}{n\left(1-(n+1)I/2+\dots\right)} \quad . \tag{2.11}$$

α approaches the trivial P/n for small I.

In order to find the total amount owed β we just multiply the installment amount α with the number of payments n:

$$\beta = \alpha n = Pn\frac{I}{1-(1+I)^{-n}} \quad . \tag{2.12}$$

Subtracting the principal P to find the total interest to be paid as γ:

$$\gamma = \alpha n - P = P\left[n\frac{I}{1-(1+I)^{-n}} - 1\right] \quad . \tag{2.13}$$

Side Note:

Note a few things straight out of this set of formulae (equations: 2.2, 2.12, 2.13):

- In order to reduce the monthly payment amount α you have several options:
  - reduce your principal P,
  - reduce the APR A or
  - increase the number of installments n which translates to a longer run time of the loan.
- To have a smaller overall cost of the loan γ you should:
  - reduce α, that is P as well as A, but
  - reduce n,

  since increasing n will generally increase γ as can be seen in figure 2.4. However, keeping the length of time constant, increasing the number of payments reduces the amount of interest paid. An illustration is given in fig. 2.2.

DERIVATIVES

It is useful to look at γ as a function of P. This is illustrated in figure 2.4. Frequently different regulatory legislation applies[8] when $\gamma < P$ than when $\gamma > P$. The latter one is scrutinized to prevent predatory lending and is all the more so watched in the short

[8] see e.g. Financial Conduct Authority regulations

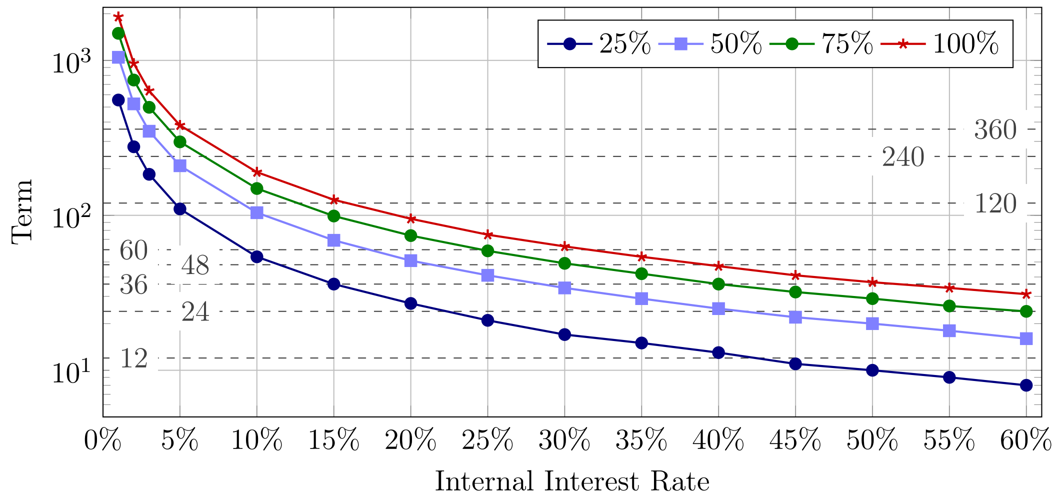


Figure 2.5: The term n as a function of A for $\kappa = 0.25, 0.5, 0.75, 1$ after equation 2.14. For example, at $A = 5\%$, a 30 year loan will result in paying the principal amount as interest, for $A = 3\%$ it will be half of the principal.

term lending space, since γ for long term products like mortgages frequently exceed P. However, rewriting equation 2.13 in terms of principal

$$\frac{\gamma}{P} = \left[n \frac{I}{1-(1+I)^{-n}} - 1 \right] = \kappa \quad , \tag{2.14}$$

provides a quick way to determine term lengths compliant with legislation as a function of I, or the internal interest rate A, respectively. An illustration is given in figure 2.5.

Limits & Approximations:

For small I, we can determine κ as

$$\lim_{I\to 0} \kappa = \frac{nI}{1-\left(1 - nI + I^2\frac{n(n+1)}{2} - \dots\right)} - 1 = \frac{1}{1 - I\frac{n+1}{2}} - 1 \quad , \tag{2.15}$$

which gives an easy to handle equation for κ as long as

$$I\frac{n+1}{2} \gg I^2\frac{(n+1)\,(n+2)}{6} \quad \to \quad I\frac{n+2}{3} \ll 1 \quad . \tag{2.16}$$

If furthermore $I(n-1)/2 \ll 1$ we can use the expansion of

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + x^5 + \dots \quad \text{for} \quad -1 < x < 1 \quad , \tag{2.17}$$

to obtain

$$\lim_{I\to 0} \kappa = \lim_{I(n+1)/2\to 0} \frac{1}{1 - I\frac{n+1}{2}} - 1 = I\frac{n+1}{2} \quad . \tag{2.18}$$

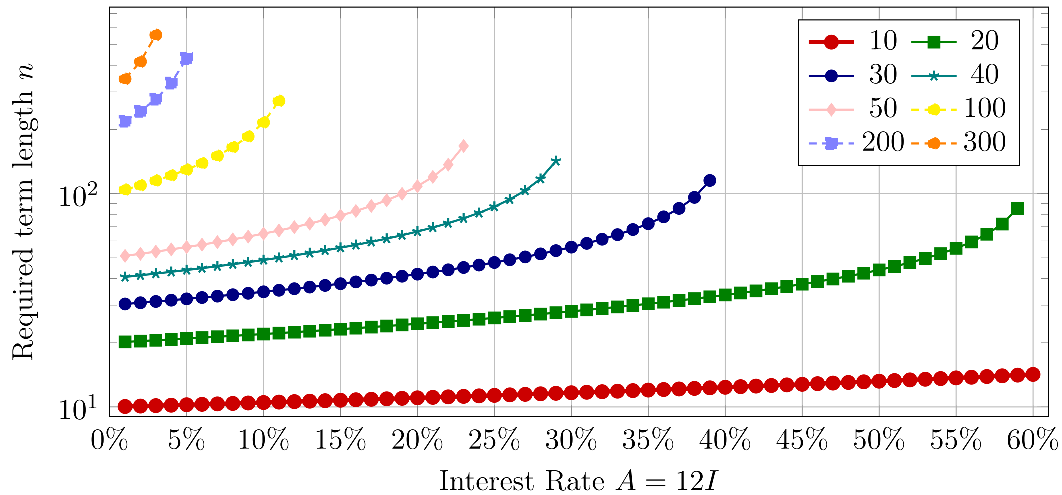


Figure 2.6: Required minimum term n as a function of interest rate $A = 12I$ for various multiples of affordable payment P/α.

Now this equation is really easy to handle and as long as we have both relations $I \ll 1$, $I(n-1)/2 \ll 1$ fulfilled reasonably well, gives a good approximation of κ for given I and n.

Example(s):

Let us assume a customer has a mortgage with a rate of $A = 3.175\%$ for a term of 10 years or $n = 120$. Equation 2.14 gives

$$\kappa = 120 \frac{0.03175/12}{1 - (1 + 0.03175/12)^{-120}} - 1 \approx 0.1685 \quad , \tag{2.19}$$

when the approximation in first order 2.15 since $0.03175/12 \approx 0.00265 \ll 1$ results in

$$\kappa \approx \frac{1}{1 - 0.03175/12 * 60.5} - 1 \approx 0.1906 \tag{2.20}$$

and consequently if we use that $0.03175 * 60.5/12 \approx 0.1601 \ll 1$, equation 2.18 gives us

$$\kappa \approx 0.03175/12 * 60.5 \approx 0.1601 \quad . \tag{2.21}$$

The quadratic correction term would roughly give us a 0.0256 on top of 0.1601 to obtain 0.1857 which is very close to the previous approximation.

Let us now assume an interest rate of $A = 10\%$ for a term of $n = 12$. We get the following three values $\kappa \approx 0.055$, $\kappa \approx 0.057$ and $\kappa \approx 0.054$.

In practice, a customer knows what rate I he qualifies for, which payment α he can afford and how much money P he needs. In that case, we can easily solve equation 2.6 for n given P, I and α:

$$n = -\frac{\ln\left(1 - \frac{PI}{\alpha}\right)}{\ln(1 + I)} \quad . \tag{2.22}$$

This equation depends on the ratio of P/α and I. An illustration is given in figure 2.6

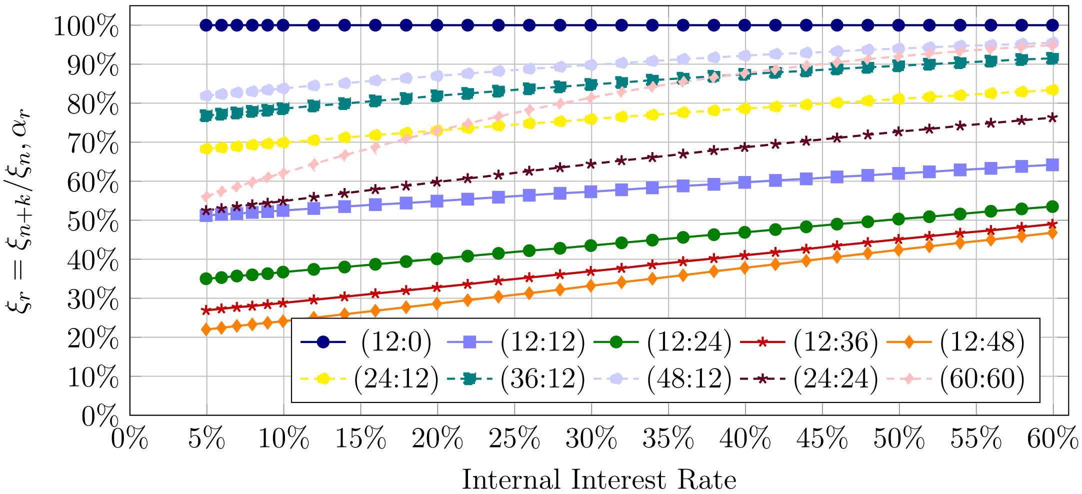


Figure 2.7: Relative fraction of payment for adding k installments to original term n as in equations 2.23 & 2.24 for $\xi_r = \xi_{n+k}/\xi_n$ which is the same as $\alpha_r = \alpha_{n+k}/\alpha$ for various combinations (n:k).

2.2.2 Dependence on Chosen Term

Given a fixed principal P and internal rate A, or I respectively, what is the effect of choosing various n onto the values of $\xi = \alpha/P$, the monthly payment relative to principal, $\tau = \beta/P$, the overall cost of the loan relative to principal and $\kappa = \gamma/P$, the overall interest paid relative to principal?

So let's add $\pm k$ installments to the original term n. We know that

$$\begin{aligned}
\xi_{n\pm k} &= \frac{\alpha_{n\pm k}}{P} = \frac{I}{1-(1+I)^{-n\mp k}} \quad , \\
\tau_{n\pm k} &= \frac{\beta_{n\pm k}}{P} = (n\pm k)\,\frac{I}{1-(1+I)^{-n\mp k}} \qquad \text{and} \\
\kappa_{n\pm k} &= \frac{\gamma_{n\pm k}}{P} = (n\pm k)\,\frac{I}{1-(1+I)^{-n\mp k}} - 1 = \tau_{n\pm k} - 1 \quad .
\end{aligned} \tag{2.23}$$

We can come up with relations of $\xi_{n\pm k}$ to the original ξ for term n, $\tau_{n\pm k}$ to the original τ as well as $\kappa_{n\pm k}$ to κ:

$$\begin{aligned}
\xi_r = \frac{\xi_{n\pm k}}{\xi} &= \frac{1-(1+I)^{-n}}{1-(1+I)^{-n\mp k}} = \frac{\alpha_{n\pm k}}{\alpha} = \alpha_r \qquad \& \\
\tau_r = \frac{\tau_{n\pm k}}{\tau} &= \frac{n\pm k}{n} \times \frac{1-(1+I)^{-n}}{1-(1+I)^{-n\mp k}} = \frac{n\pm k}{n}\xi_r \qquad \& \\
\kappa_r = \frac{\kappa_{n\pm k}}{\kappa} &= \frac{(n\pm k)\,\frac{I}{1-(1+I)^{-n\mp k}} - 1}{n\frac{I}{1-(1+I)^{-n}} - 1} \\
&= \frac{1-(1+I)^{-n}}{1-(1+I)^{-n\mp k}} \times \frac{(n\mp k)\,I - 1 + (1+I)^{-n\mp k}}{nI - 1 + (1+I)^{-n}} \\
&= \xi_r \frac{(n\mp k)\,I - 1 + (1+I)^{-n\mp k}}{nI - 1 + (1+I)^{-n}} \quad .
\end{aligned} \tag{2.24}$$

Apparently $\xi_{n+k} < \xi$, or $\xi_r < 1$ for all positive k. A graphical representation for various examples is given in figure 2.7. τ_r on the other hand is larger than 1 for all positive k. Obviously τ_r is larger than ξ_r by a factor of $(n+k)/n$.

Limits & Approximations:

Using equation 2.10 we obtain

$$\begin{aligned}
\lim_{I\to 0} \xi_r &= \frac{n}{n \pm k} \quad , \\
\lim_{I\to 0} \tau_r &= \frac{n \pm k}{n} \times \frac{I \times n - I^2 n(n+1)/2}{I \times (n \pm k) - I^2 (n \pm k)(n \pm k + 1)/2} \\
&= \frac{2 - I(n+1)}{2 - I(n \pm k + 1)} \qquad \text{and} \\
\lim_{I\to 0} \kappa_r &= \frac{n}{n \pm k} \times \frac{I\,(n \pm k) - 1 + \lim_{I\to 0}(1+I)^{-n\mp k}}{nI - 1 + \lim_{I\to 0}(1+I)^{-n}} \\
&= \frac{n}{n \pm k} \times \frac{I^2}{I^2} \times \frac{(n \pm k)}{n} \times \frac{(n \pm k + 1)/2}{(n+1)/2} \\
&= \frac{n \pm k + 1}{n+1} \quad . \qquad (2.25)
\end{aligned}$$

The limits for ξ_r are clearly showing in figure 2.7 as is the effect of doubling term. With the original n being smaller, the effect is more pronounced and vanishes quickly as interest rates increase. The limit for $n \to \infty$ of ξ_r is 1.

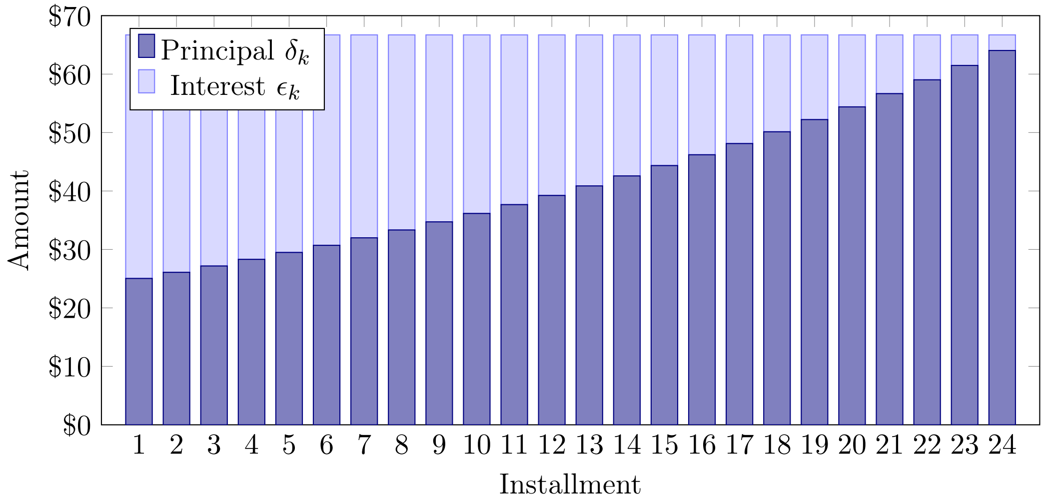


Figure 2.8: δ_k and ϵ_k for a loan with with $n = 24$, $P = \$1000$ and $A = 50\%$, $k = 1 \ldots n$.

2.2.3 Installment Level Interest and Principal

Breaking out the amount of interest ϵ_k for any given payment α_k of installment k we simply multiply the remaining principal of the previous payment P_{k-1} with I and obtain:

$$\epsilon_k = P_{k-1} I \tag{2.26}$$

Obviously the reduction of principal δ_k is $\delta_k = \alpha_k - \epsilon_k$ and $P_k = P_{k-1} - \delta_k$.

For those of us who prefer an explicit formula to iterative computations, we can then compute the remaining principal P_k by plugging equation 2.6 into equation 2.4 as

$$\begin{aligned} P_k &= P(1+I)^k - P\left[(1+I)^k - 1\right] \times \left[\frac{(1+I)^n}{(1+I)^n - 1}\right] \\ &= P\frac{(1+1)^n - (1+I)^k}{(1+I)^n - 1} = P\left[1 - \frac{(1+I)^k - 1}{(1+I)^n - 1}\right]. \end{aligned} \tag{2.27}$$

We can now very easily determine

$$\epsilon_k = P_{k-1} I = P \times I \frac{(1+1)^n - (1+I)^{k-1}}{(1+I)^n - 1} = P \times I\left[1 - \frac{(1+I)^{k-1} - 1}{(1+I)^n - 1}\right] \tag{2.28}$$

as well as

$$\begin{aligned} \delta_k &= P_{k-1} - P_k = P\left[1 - \frac{(1+I)^{k-1} - 1}{(1+I)^n - 1}\right] - P\left[1 - \frac{(1+I)^k - 1}{(1+I)^n - 1}\right] \\ &= P\frac{(1+I)^k - (1+I)^{k-1}}{(1+I)^n - 1} = P\frac{(1+I)^{k-1}(1+I-1)}{(1+I)^n - 1} = PI\frac{(1+I)^{k-1}}{(1+I)^n - 1} \end{aligned} \tag{2.29}$$

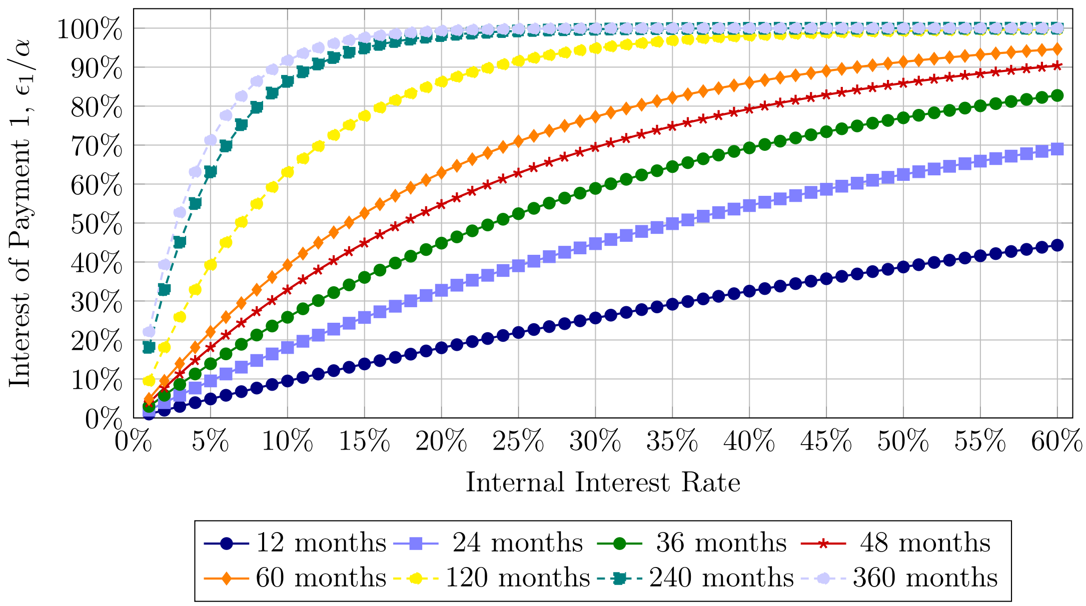


Figure 2.9: The share of interest of the first payment amount ϵ_1 for various interest rates (x-axis) and terms.

Consistency Check:

Let us quickly check the sanity of the computation!

$$\begin{aligned}\delta_k + \epsilon_k &= P \times I \frac{(1+I)^{k-1}}{(1+I)^n - 1} + P \times I \left[1 - \frac{(1+I)^{k-1} - 1}{(1+I)^n - 1}\right] \\ &= P \times I \left[\frac{(1+I)^{k-1} + (1+I)^n - 1 - (1+I)^{k-1} + 1}{(1+I)^n - 1}\right] \\ &= \frac{P \times I}{1 - (1+I)^{-n}} = \alpha \quad . \end{aligned} \tag{2.30}$$

How much principal δ_k is paid relative to the payment amount α at payment k ?

$$\begin{aligned}\frac{\delta_k}{\alpha} &= P \times I \frac{(1+I)^{k-1}}{(1+I)^n - 1} \times \frac{1 - (1+I)^{-n}}{P \times I} \\ &= \frac{(1+I)^{k-1}}{(1+I)^n} \times \frac{1 - (1+I)^{-n}}{1 - (1+I)^{-n}} = (1+I)^{k-1-n} \end{aligned} \tag{2.31}$$

and similarly

$$\begin{aligned}\frac{\epsilon_k}{\alpha} &= P \times I \left[1 - \frac{(1+I)^{k-1} - 1}{(1+I)^n - 1}\right] \times \frac{1 - (1+I)^{-n}}{P \times I} \\ &= \frac{(1+I)^n - 1 - (1+I)^{k-1} + 1}{(1+I)^n} = 1 - (1+I)^{k-1-n} = 1 - \frac{\delta_k}{\alpha} \quad . \end{aligned} \tag{2.32}$$

The share of interest of the first payment is given in figure 2.9. A depiction of how the share of interest evolves with A is given in figure 2.10 for $n = 12$ and $k = 1, \ldots, n$.

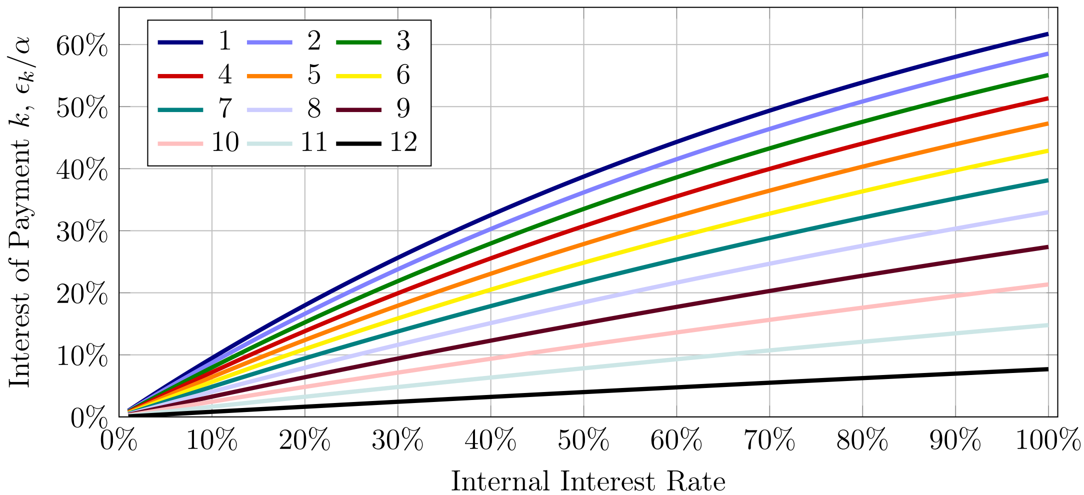


Figure 2.10: The share of interest of each payment amount ϵ_k/α for various interest rates (x-axis) and $k = 1, \ldots, 12$ at a term length of $n = 12$. $k = 1$ is the topmost curve, $k = 12$ the lowest one. As shown, the share of interest on each payment is reduced as part of the amortization, which corresponds to a rising share of principal in each payment vs. the previous one.

Equation 2.27 immediately leads to the amount of interest paid up till that point η_k when we put the remaining amount owed to $P - P_k$

$$\begin{aligned} \eta_k &= \alpha k - (P - P_k) \\ &= P\frac{kI}{1-(1+I)^{-n}} - P + P\left[1 - \frac{(1+I)^k - 1}{(1+I)^n - 1}\right] \\ &= P\left[\frac{kI + 1 - (1+I)^{k-n}}{1-(1+I)^{-n}} - 1\right] = P\frac{kI + (1+I)^{-n} - (1+I)^{k-n}}{1-(1+I)^{-n}} \\ &= P\frac{kI\,(1+I)^n - (1+I)^k + 1}{(1+I)^n - 1} \quad . \end{aligned} \tag{2.33}$$

If we determine $\theta_k = \eta_k/P$ we obtain

$$\theta_k = \frac{kI\,(1+I)^n - (1+I)^k + 1}{(1+I)^n - 1} \quad . \tag{2.34}$$

θ_k depends on three parameters I, k and n. In any analysis of one of them, it makes sense to group the other two to get a more accurate read.

Example(s):

Table 2.2 shows an example of an amortization plan for a 24 month loan of \$1,000.00 at an APR of 50%. If the customer meets all his payments in time the total interest paid is \$601.06 which corresponds to ≈30% per annum as opposed to the 50% one could naively expect.

Table 2.2: Amortization table for a loan with $A = 50\%$, $P = \$1000$ and $n = 24$. It shows for each installment k the payment amount α, the interest paid ϵ_k, the principal δ_k, its share δ_k/α, the relation of interest to principal ϵ_k/δ_k, the payments made $k \times \alpha$, the overall interest paid η_k, the overall interest paid relative to loan amount θ_k and the outstanding principal P_k.

k	α (\$)	ϵ_k (\$)	δ_k (\$)	δ_k/α	ϵ_k/δ_k	$k \times \alpha$	η_k (\$)	θ_k	P_k (\$)
0	0.00	0.00	0.00	0.000	0.000	0.00	0.00	0.00	1,000.00
1	66.71	41.67	25.04	0.375	1.664	66.71	41.67	0.04	974.96
2	66.71	40.62	26.09	0.391	1.557	133.42	82.29	0.08	948.87
3	66.71	39.54	27.17	0.407	1.455	200.13	121.83	0.12	921.69
4	66.71	38.40	28.31	0.424	1.357	266.84	160.23	0.16	893.39
5	66.71	37.22	29.49	0.442	1.262	333.55	197.45	0.20	863.90
6	66.71	36.00	30.71	0.460	1.172	400.26	233.45	0.23	833.19
7	66.71	34.72	31.99	0.480	1.085	466.98	268.17	0.27	801.19
8	66.71	33.38	33.33	0.500	1.002	533.69	301.55	0.30	767.86
9	66.71	31.99	34.72	0.520	0.922	600.40	333.54	0.33	733.15
10	66.71	30.55	36.16	0.542	0.845	667.11	364.09	0.36	696.98
11	66.71	29.04	37.67	0.565	0.771	733.82	393.13	0.39	659.31
12	66.71	27.47	39.24	0.588	0.700	800.53	420.60	0.42	620.07
13	66.71	25.84	40.87	0.613	0.632	867.24	446.44	0.45	579.20
14	66.71	24.13	42.58	0.638	0.567	933.95	470.57	0.47	536.62
15	66.71	22.36	44.35	0.665	0.504	1,000.66	492.93	0.49	492.27
16	66.71	20.51	46.20	0.693	0.444	1,067.37	513.44	0.51	446.07
17	66.71	18.59	48.12	0.721	0.386	1,134.08	532.03	0.53	397.95
18	66.71	16.58	50.13	0.751	0.331	1,200.79	548.61	0.55	347.82
19	66.71	14.49	52.22	0.783	0.278	1,267.50	563.10	0.56	295.60
20	66.71	12.32	54.39	0.815	0.226	1,334.22	575.42	0.58	241.20
21	66.71	10.05	56.66	0.849	0.177	1,400.93	585.47	0.59	184.54
22	66.71	7.69	59.02	0.885	0.130	1,467.64	593.16	0.59	125.52
23	66.71	5.23	61.48	0.922	0.085	1,534.35	598.39	0.60	64.04
24	66.71	2.67	64.04	0.960	0.042	1,601.06	601.06	0.60	0.00

Side Note:

Equation 2.31 has a few properties that are worth mentioning.

If we look at δ_k/α, how does it relate to δ_{k+1}/α when $0 < k < k+1 \leq n$?

$$\frac{\delta_{k+1}}{\alpha} = (1+I)^{k+1-1-n} = (1+I)^{k-1-n}(1+I) = (1+I) \times \frac{\delta_k}{\alpha} \quad . \tag{2.35}$$

Since $(1+I) > 1$ for all positive I, the amount of principal paid down is increasing for each installment which is a fundamental property of amortizing loans. Similarly, the relative amount of interest paid per installment will go down over time.

A customer or regulator could also want no installment payment to consist of more than x percent interest. Using equation 2.35, we know that the share of principal in any installment is minimal for the first installment due. We can

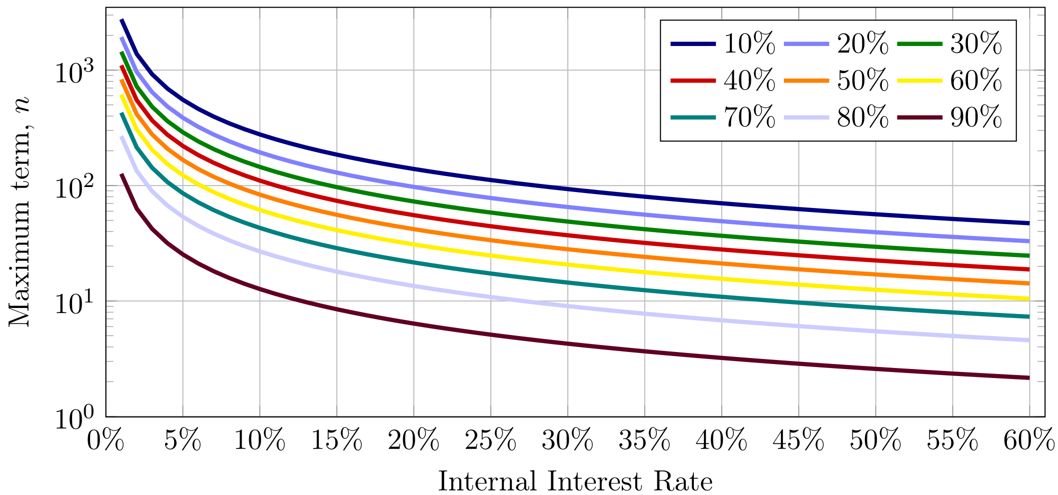


Figure 2.11: Maximum term ensuring the principal makes up the given share of the installment payment for $\delta_1/\alpha = 10\% \ldots 90\%$. As an example, the orange curve puts the principal at 50% of the installment amount, we can easily find that for e.g. an interest rate of 20% the term n cannot exceed 41 installments to hold this relation as given in equation 2.37.

therefore use 2.31 and $k = 1$ to come up with the minimum δ_k/α via

$$\frac{\delta_k}{\alpha}|_{\min} = \frac{\delta_1}{\alpha} = (1+I)^{1-1-n} = (1+I)^{-n} \quad , \tag{2.36}$$

from which we can easily deduce the maximum term for which this relation exceeds a given value of x:

$$n \leq -\frac{\ln x}{\ln(1+I)} \quad . \tag{2.37}$$

This relation is depicted in figure 2.11. Similarly, if we fix the term, we can come up with a relation for the maximum allowable interest rate I:

$$I \leq x^{-1/n} - 1 \quad . \tag{2.38}$$

The result is illustrated in figure 2.12.

Equation 2.36 allows us to express e.g. κ in dependence of n, I and δ_1. In other words, we can determine the interest relative to principal paid over the lifetime of the loan from the share of principal in the first payment, the term and interest rate and vice versa.

$$\kappa = n\left[\frac{I}{1-\delta_1} - 1\right] \quad \text{or} \quad \delta_1 = 1 - \frac{nI}{\kappa + n} \tag{2.39}$$

respectively.

Another interesting relation is - of course - the ratio between interest and principal in each installment k:

$$\frac{\epsilon_k/\alpha}{\delta_k/\alpha} = \frac{\epsilon_k}{\delta_k} = \frac{1-(1+I)^{k-1-n}}{(1+I)^{k-1-n}} = \frac{1}{(1+I)^{k-1-n}} - 1 \quad . \tag{2.40}$$

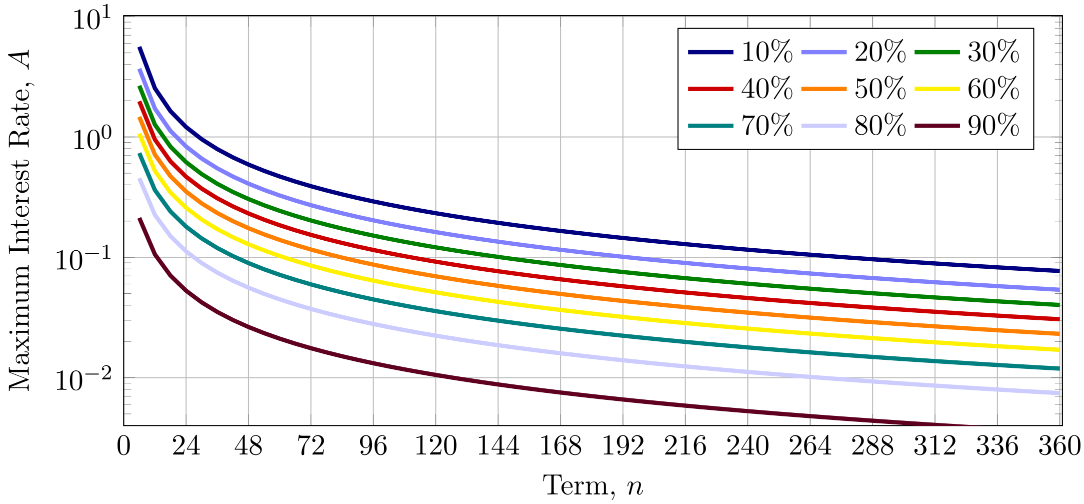


Figure 2.12: Maximum interest rate ensuring the principal makes up the given share of the installment payment for $\delta_1/\alpha = 10\% \ldots 90\%$. As an example, the orange curve puts the principal at 50% of the installment amount, we can easily find that for e.g. a term of 42 months the interest rate cannot exceed 20% to hold this relation as given in equation 2.38.

Consistency Check:

If a customer paid the first k installments and would be offered the same interest rate A, what would the payment amount α_{n-k} be for a recomputed loan with term $n-k$?

$$\begin{aligned}
\alpha_{n-k} &= \frac{P_k I}{1-(1+I)^{-n+k}} = P_k \frac{I\,(1+I)^{n-k}}{(1+I)^{n-k}-1} \\
&= P\left[1-\frac{(1+I)^k-1}{(1+I)^n-1}\right] \times \frac{I\,(1+I)^{n-k}}{(1+I)^{n-k}-1} \\
&= P\frac{(1+I)^n-(1+I)^k}{(1+I)^n-1} \times \frac{I\,(1+I)^{n-k}}{(1+I)^{n-k}-1} \\
&= P\left[\frac{(1+I)^n}{(1+I)^n-1}\right] \times I\frac{(1+I)^{n-k}-1}{(1+I)^{n-k}-1} \\
&= PI\left[\frac{(1+I)^n}{(1+I)^n-1}\right] = \alpha
\end{aligned} \tag{2.41}$$

The payment amount does not change given the same interest rate, remaining principal and term. This is a fundamental property of installment loans. If you qualified for a better interest rate though, you will save. This is the whole idea behind refinancing customers with loans in good standing.

Example(s):

Examples, graph this big time by I and n, example with mortgage

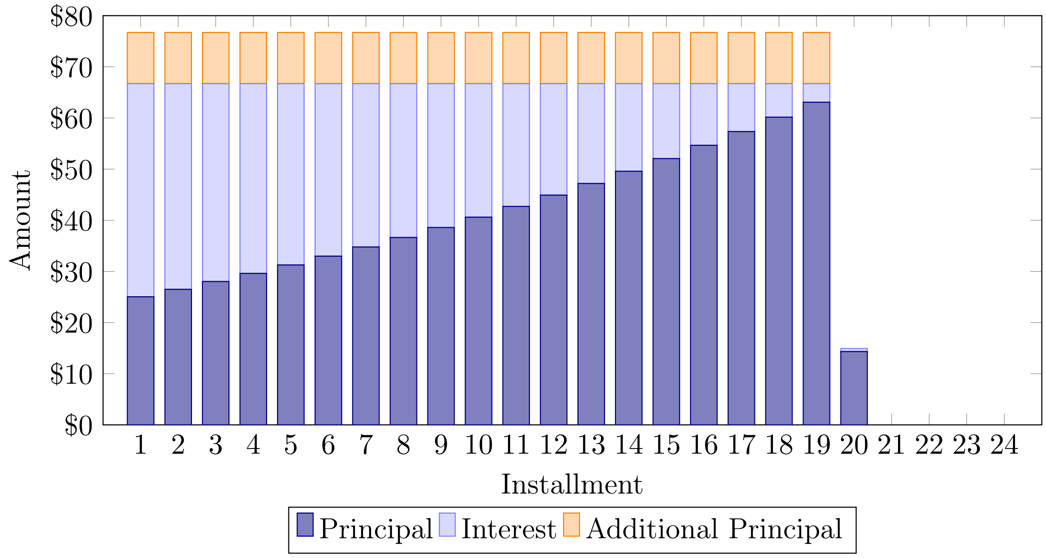


Figure 2.13: Shares of interest and principal for the loan discussed above, 24 months, \$1000, 50% APR. Customer decided to overpay each installment by \$10 or 15% of his designated installment amount.

2.2.4 Overpaying Principal & Early Payoff

A frequently encountered case is the customer paying down additional principal by overpaying his monthly installment. Let the overpayment $\chi \geq 0$ be given as a fraction of the designated installment amount α, so the total monthly payment is $\alpha(1+\chi) = \alpha\rho$ with $\rho \geq 1$. We then find how many payments p have to be made by following a similar method as in equations 2.2 to 2.6

$$\alpha(1+\chi) = \alpha\rho = P\frac{I}{1-(1+I)^{-p}} \quad , \tag{2.42}$$

which we can use to determine p

$$p = -\frac{\ln\left[1-\frac{PI}{\alpha\rho}\right]}{\ln(1+I)} = -\frac{\ln\left[1-\frac{1-(1+I)^{-n}}{1+\chi}\right]}{\ln(1+I)} \tag{2.43}$$

This again, is independent of the principal amount. So for A=50%, n=24, we find a $\chi = 0.5$ over payment to result in a term of $p = 14$, for $\chi = 1$ we get $p = 10$, with the last payment representing a fractional amount of the original one α.

To find out how much less a customer paid, we approximate[9] the amount of savings s by

$$s \approx n\alpha - p\alpha\rho = \alpha(n - p\rho) \quad , \tag{2.44}$$

which of course is all interest. The above equation could easily be written explicitly, but it turns lengthy. Since α is positive at all times, we yield positive savings as long as $n > p\rho$ or $p < n/\rho$ which is a good exercise to mathematically prove that this is always the case. Intuitively we would expect such behavior. On the numbers it is interesting to notice that the number of payments reduces to fewer than n/ρ.

[9] Since the fractional payment accrues interest, the amount will be off by a few cents. Maybe I add this calculation at some later stage.

Example(s):

An illustration is given in figure 2.13 for an example of a \$10 or 15% overpayment of the monthly installment amount. Payment 20 is the last payment for a smaller amount. the savings in cost of loan can be seen in figure 2.8. There are four installments saved in this configuration, or in other words, the term has shortened by 17%. The amount of interest saved is \$128.62 or more than 21% of the original loan cost. Both of these savings exceed the 15% value of the overpayment. In terms of original installment amount the total savings over the course of the loan is roughly 2 installment payments. If computed by formula 2.44 the amount of savings s would have been $s \approx \$128.74$.

Graphical representation in terms of n and A, also installment plan for an example, explicit expression calculation on how much interest was saved

Compute the actual interest paid divided by year in percent of principal, compare to original values, relate effective internal interest rate to rate before as dependent on χ

Analysis of whether or not it makes sense to take out higher than desired principal, then pay back excess amount in first payment vs. taking out lower amount or shorter term

Mention difference installment loans shortening term vs. LOC or CC reducing min loan amount when payment more than intended.

Side Note:

All calculations above are notwithstanding missed payments, grace periods, courtesies, late payments and fees for such. These should be collected as additional payments on top of any given installment. If we cannot get the additional fees from the customer, we maintain the installment structure and values, but take out fees before interest and then principal. It is in this case possible to not reduce the principal in a given installment. Additionally the time a payment is late accrues additional interest. We do not automatically add additional installments at the end of term, but will collect outstanding money. The loan will not be considered paid off until that balance is paid event though all installment amounts have been paid. The earlier a missed payment occurs, the more damage it does.

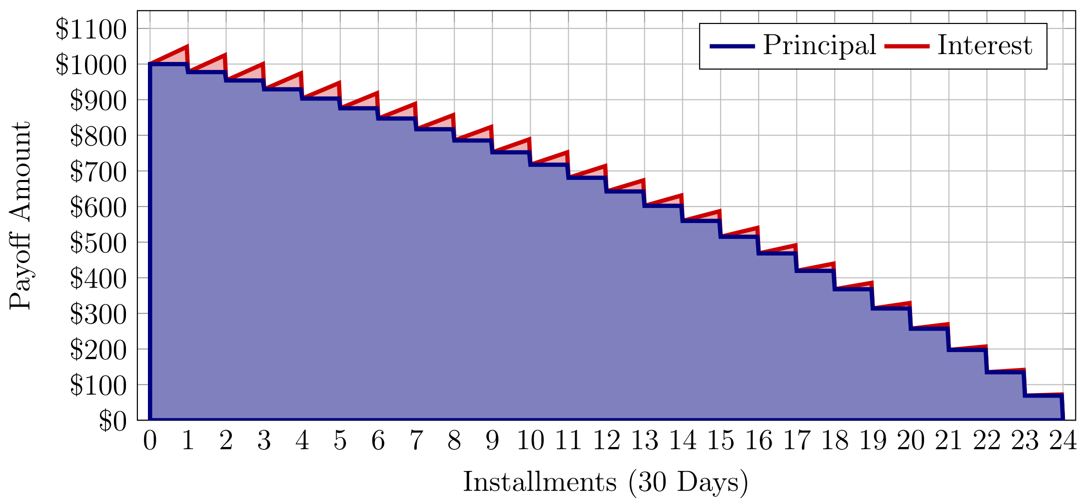


Figure 2.14: Payoff

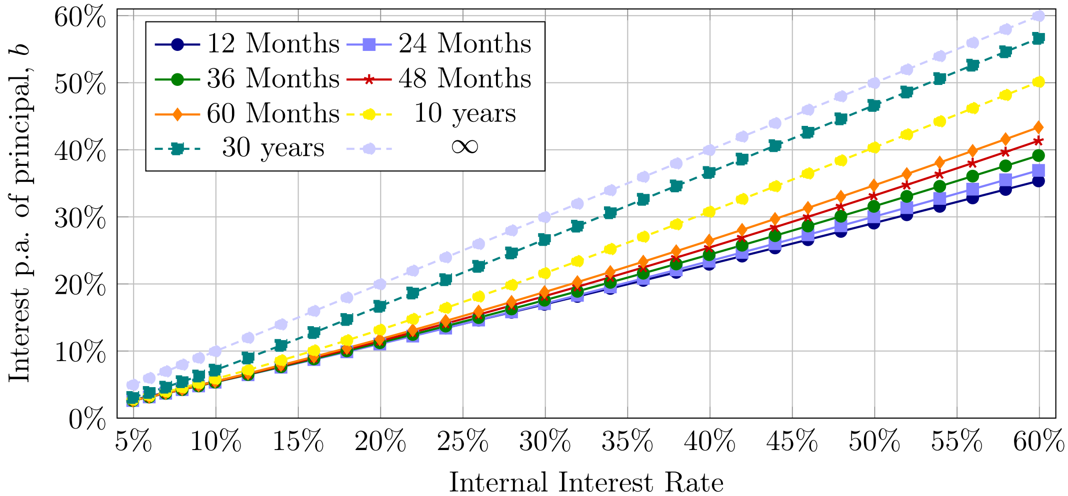


Figure 2.15: Interest paid per year in percent of principal, b, vs. internal interest rate A at various terms lengths. For $n \to \infty$ we obtain the identity function, all finite terms result in interest paid per year of less that what multiplying the internal rate A by the number of years $n/12$ would suggest.

2.2.5 Overall Cost of Borrowing vs. Interest Rate and Term

For amortizing loans[10], A applied to the principal P over the length of the term n in years n/m is **not equal to and usually much higher than** the total interest to be paid γ divided by the number of years which we shall refer to as b.

$$\begin{aligned} A \neq b &= \gamma / \left[P \frac{n}{m} \right] \\ &= \frac{m}{n} \left[n \frac{I}{1-(1+I)^{-n}} - 1 \right] \quad . \end{aligned} \tag{2.45}$$

This percentage does not depend on the principal amount.

Limits & Approximations:

For $n \to \infty$ the value for the interest paid per year approaches A as illustrated in figure 2.15.

$$\begin{aligned} \lim_{n\to\infty} b &= \lim_{n\to\infty} \gamma / \left[P \frac{n}{m} \right] \\ &= \lim_{n\to\infty} \frac{m}{n} \left[n \frac{I}{1-(1+I)^{-n}} - 1 \right] \\ &= mI \\ &= A \quad . \end{aligned} \tag{2.46}$$

Let us for a moment consider the total interest paid divided by the number of years

[10] Principal is reduced with each payment.

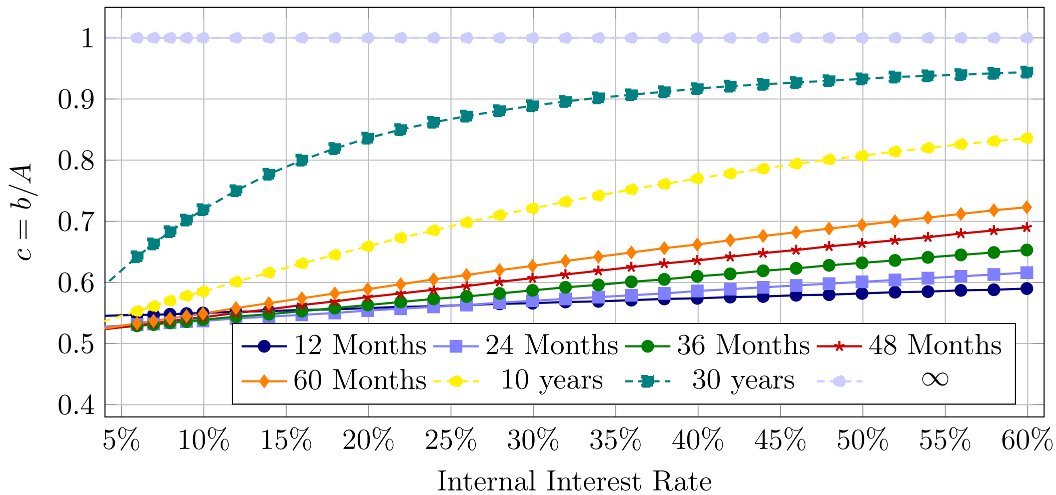


Figure 2.16: Interest paid per year relative to the internal rate $c = b/A$.

as a percentage of internal interest rate $c = b/A$:

$$\begin{aligned} c &= b/A \\ &= \frac{m}{n}\left[n\frac{I}{1-(1+I)^{-n}} - 1\right] \times \frac{1}{I \times m} \\ &= \frac{1}{n \times I}\left[n\frac{I}{1-(1+I)^{-n}} - 1\right] \quad , \end{aligned} \tag{2.47}$$

which is independent of m or how many payments there are in a year. It only depends on the number of payments n and the interest rate I. An illustration is given in figure 2.16.

Limits & Approximations:

Apparently for large n it approaches

$$\begin{aligned} \lim_{n\to\infty} c &= \lim_{n\to\infty} \frac{1}{n \times I}\left[n\frac{I}{1-(1+I)^{-n}} - 1\right] \\ &= \lim_{n\to\infty}\left[1 - \frac{1}{n \times I}\right] \\ &= 1 \quad . \end{aligned} \tag{2.48}$$

Taking a closer look at equation 2.47 reveals that there are two competing terms in the formula. This would suggest that possibly an extremum for c exists. As shown in figure 2.17 there is a sweet spot for the term n as a function of I. There is a value of n that assumes a minimum of c for each I. In other words, choosing term carefully can reduce the effective interest paid per unit time. Let's figure out a functional relationship describing the optimum term in this respect.

Limits & Approximations:

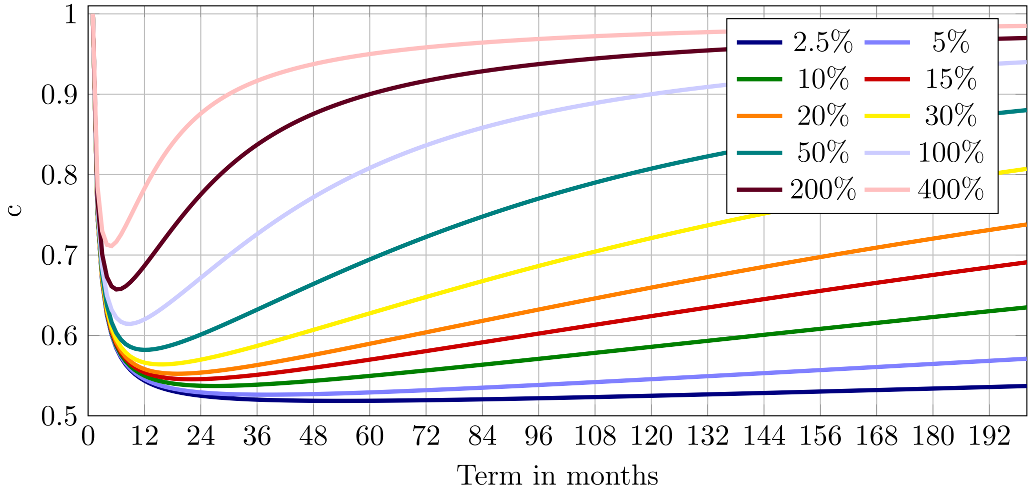


Figure 2.17: The interest per year c for different values of $A = 12I$ as a function of term n.

Using the Taylor expansion in equation 2.10 we obtain[a]

$$
\begin{aligned}
\lim_{I\to 0} c &= \lim_{I\to 0} \frac{1}{n\times I}\left[n\frac{I}{1-(1+I)^{-n}} - 1\right] \\
&= \lim_{I\to 0} \frac{1}{n\times I}\frac{n\times I - 1 + 1 - I\times n + I^2\frac{n(n+1)}{2} + \ldots}{1 - 1 + I\times n + \ldots} \\
&= \frac{I^2}{2n^2I^2}n(n+1) = \frac{n(n+1)}{2n^2} = \frac{n+1}{2n} \\
&= \frac{1}{2} + \frac{1}{2n} \quad . \qquad (2.49)
\end{aligned}
$$

This implicates that for small I and large n the value of c goes asymptotically to $c \to \frac{1}{2}$. For the smallest n, a single installment, $n = 1$, the value obviously is $c = 1$. For $I \to \infty$ we get

$$
\begin{aligned}
\lim_{I\to \infty} c &= \lim_{I\to \infty} \frac{1}{n\times I}\left[n\frac{I}{1-(1+I)^{-n}} - 1\right] \\
&= 1 - \frac{1}{I\times n} \\
&= 1 \quad , \qquad (2.50)
\end{aligned}
$$

independent of the value of n. It is fairly straightforward to prove that $\frac{1}{2} < c \leq 1$ for all I, n.

[a] We expand to the first term that doesn't cancel for both numerator and denominator. That means in this case expanding the numerator to second order terms, the denominator to first order term and ignore the rest for being irrelevant.

compute derivative v I and derivative v n for maximum impact and value or lim of $I \to 0$, show function rises with n and I

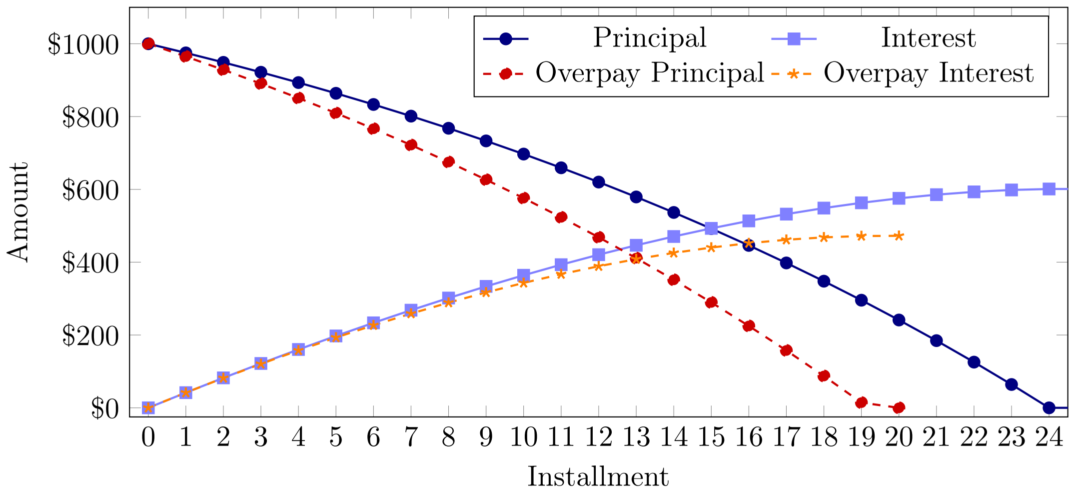


Figure 2.18: Remaining Principal and Interest collected (solid curves) for the loan discussed above, 24 months, $1000, 50% APR. As we can see very clearly in this graph, at installment 15, the collected interest exceeds the remaining principal. Additional curves (dashed) added for an overpayment of $10 or 15% per installment.

2.3 Business Considerations

2.3.1 Individual Loan Earnings Analysis

As a lender, the money we make comes from the interest part and possibly fees and extra charges. As **AVANT** strives for a fair and transparent process and customer experience, there are no additional fees or charges on the happy path other than interest. This means, that our margin per loan consists entirely of the total interest γ. Looking at equation 2.13 we like - as long as risk of default is zero - to have the highest A or I respectively, large principals P and long terms n.

The business task at hand is of course, to balance risk of default with attractive market rates and therewith maximize the revenue. The determination of that risk as accurately and aggressively as possible is where the money is.

So, one important question to answer is, when (after which installment k) do we start making money on any given loan (get more in payments than we gave out as principal). This can be easily calculated as

$$k \geq P/\alpha = \frac{1-(1+I)^{-n}}{I}. \tag{2.51}$$

We round the value up to the nearest integer, of course. k will be smaller the higher I is and smaller n and **does not depend on** P. However, in relation to the length of term, k/n we reach a profit sooner the longer the loan runs, ergo n increases.

Example(s):

In the above example, the break even is after Installment $k = 15$ or 63% of the length of term. Looking at the customer overpaying his installments with additional principal we reach a break even point sooner, but make less money overall. In the example we were to obtain $k = 13$.

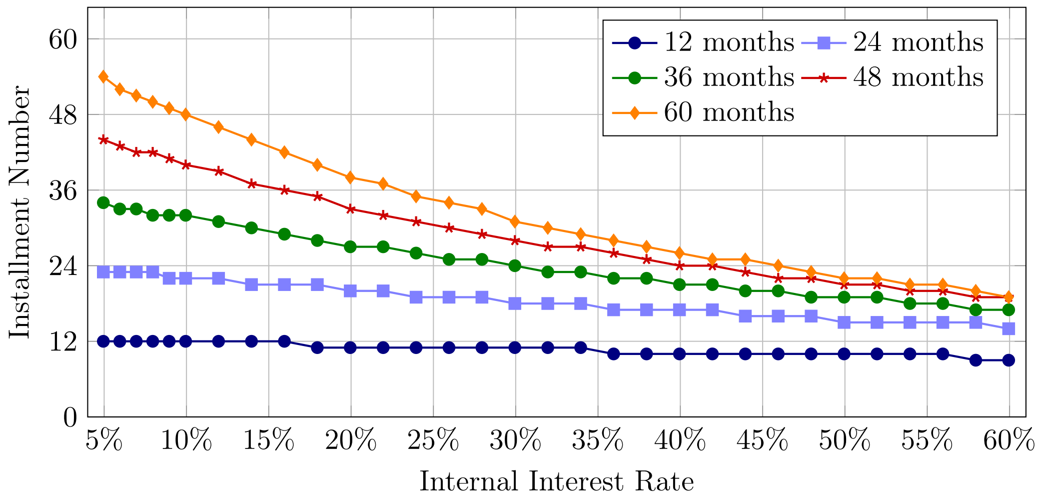


Figure 2.19: Breakeven installment number versus term according to equation 2.51 evolution with term n.

It is also worth noting that the interest paid per installment is proportional to the outstanding principal. That has a few important consequences, the amount of profit we make is the highest on the first few installments. Also, if a customer chooses to pay down his loan immediately after taking it out we make no money at all or very little[11] depending on state.

How do costs incurred play into the benefit analysis? At this point generally we care about the cost we paid to fund the customer's loan, the cost of capital and the cost to service the loan. Let us ignore the cost of capital for now as well as the exact composition of the cost of funding and servicing the loan [12]. Assuming there is a fixed amount μ that we paid in addition to the principal. Obviously this only makes sense (or even theoretical profit) when $P + \mu$ is smaller than $n \times \alpha$. Equation 2.51 will transform into

$$\begin{aligned} k \times \alpha &\geq P + \mu \\ k &\geq \frac{1-(1+I)^{-n}}{I} \times (1 + \frac{\mu}{P}) \quad . \end{aligned} \tag{2.52}$$

How much can μ/P be for given n, I to break even (make no money). Apparently then $P + \mu = n \times \alpha$ which dictates

$$\frac{\mu}{P} \leq \frac{n \times \alpha}{P} - 1 = \frac{\gamma}{P} \tag{2.53}$$

choose profit in £of ω then determine max μ.

Paying off the loan after a certain, long time span during the lifetime of the loan is the more desirable to us the smaller the number of installments remaining. It is desirable overall though, to get our money back regardless rather than not at all.

[11] Which is why other lenders charge an early payoff fee.

[12] See section bla.

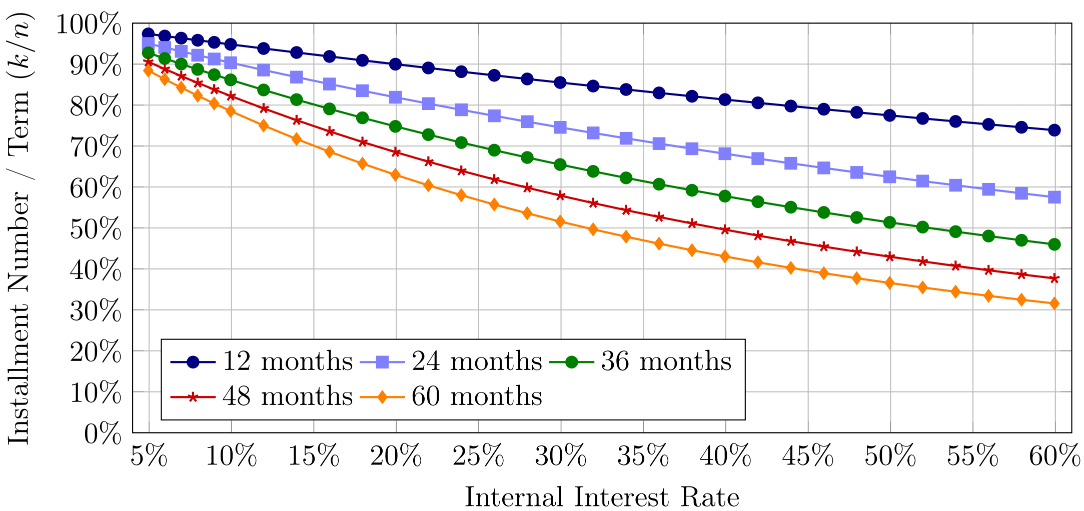


Figure 2.20: Breakeven installment number relative to full term term according to equation 2.51 evolution with term n.

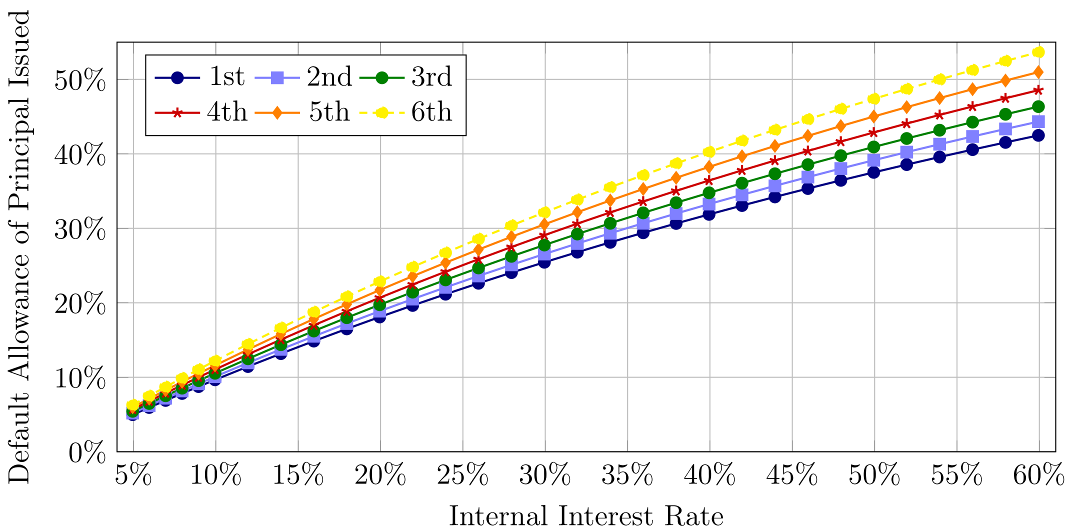


Figure 2.21: Default allowance by installment provided there are no other costs such as prepayment, ops, marketing and previous defaults. This is unrealistic, of course, but gives a feel how equation 2.55 evolves with the parameter k=1,...,6 for term n=24 months.

2.3.2 Groupings of Loans

Provided all other parameters stay equal within a class of loans, how many fully paid back loans j_k do we need to make up for potential losses due to defaults at payment k?[13] This can easily be answered by relating total interest γ to the amount of principal P minus the amount of payments received after k-1 installments $\alpha(k-1)$. Plugging in equations 2.6 and 2.13, we obtain:

$$
\begin{aligned}
j_k \geq \frac{P-\alpha(k-1)}{\gamma} &= \frac{P - P(k-1)\frac{I}{1-(1+I)^{-n}}}{P\left[n\frac{I}{1-(1+I)^{-n}} - 1\right]} \\
&= \frac{1-(1+I)^{-n}-(k-1)I}{nI-1+(1+I)^{-n}} \\
&= \frac{\left[1-I(k-1)\right](1+I)^n-1}{(nI-1)(1+I)^n+1} \qquad , \qquad (2.54)
\end{aligned}
$$

which is also beautifully independent of P.

The default rate d_k that we need to be below of to break even follows as we need j_k loans to be good - that is paid in full - for every 1 loan defaulting at installment k:

$$
\begin{aligned}
d_k \leq \frac{1}{j_k+1} &= \frac{1}{\frac{[1-I(k-1)](1+I)^n-1}{(nI-1)(1+I)^n+1}+1} \\
&= \frac{(nI-1)(1+I)^n+1}{\left[1-I(k-1)\right](1+I)^n-1+(nI-1)(1+I)^n+1} \\
&= \frac{(nI-1)(1+I)^n+1}{I(n+1-k)(1+I)^n} \qquad . \qquad (2.55)
\end{aligned}
$$

[13] Disregarding COGS, of course, so this number represents the very minimum and serves only as an illustration of business principles

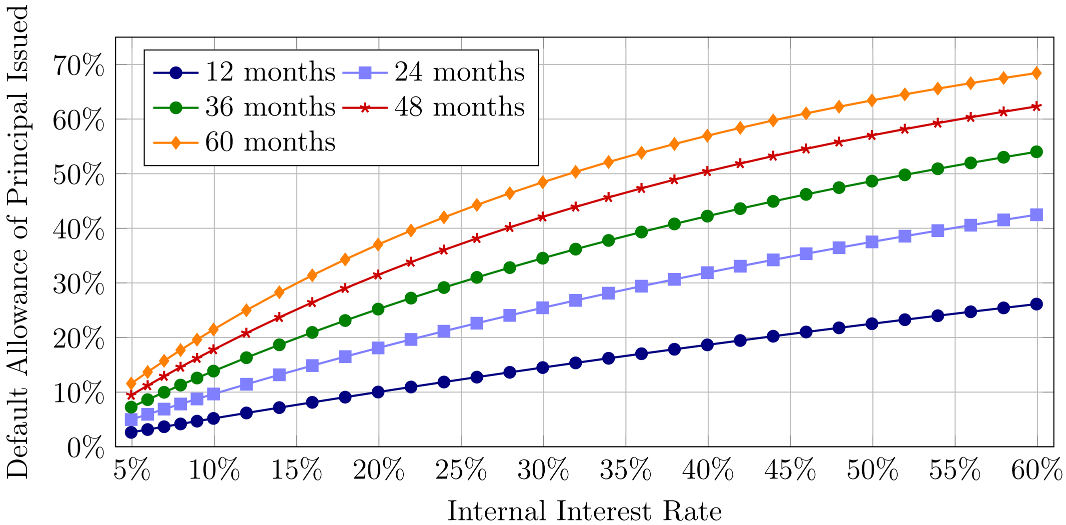


Figure 2.22: Default allowance of first installment by term provided there are no other costs such as prepayment, ops, marketing and previous defaults. This is unrealistic, of course, but gives a feel how equation 2.55 evolves with term n.

This expression is a very interesting one. It only depends on three parameters, n, I and k. The first thing which that pops is that the higher k is, meaning the later the default in the payment process, the higher of a default rate we can theoretically afford, or, in other words, the interest rate we need to charge can be decreased with k. That makes a lot of sense.

It also increases with number of payments n which relates directly to the length of term as well as with I which in turn directly relates to A. So if we do move I and n in different directions, the effects of either may balance in a specific fashion as shown above, but as a rule of thumb, we can afford to go lower in A if at the same time we increase n if we wanted to preserve d_k. This also means that we would want to charge more APR the shorter the term is.

Example(s):

Let us look at a one installment loan, $n = 1, k = 1$. Equation 2.55 then converts to:

$$\begin{aligned} d_{(1)} &= \frac{(I-1) \times (1+I) + 1}{I \times (1+I)} \\ &= \frac{I}{I+1} \quad . \end{aligned} \tag{2.56}$$

We can transform this into an expression for I as a function of $d_{(1)}$:

$$I = \frac{d_{(1)}}{1 - d_{(1)}}. \tag{2.57}$$

Apparently that I is larger than $d_{(1)}$ in this case. Using $A = m \times I$ we can get the formal internal interest rate easily.

Let us now consider $n = 2$ apparently there are two terms one for $k = 1$, another for $k = 2$, each assuming we are not losing any payments on the respective other. Using equation 2.55 we find:

$$\begin{aligned} d_k &= \frac{(2I-1)\times(1+I)^2+1}{I(3-k)\times(1+I)^2} \\ &= \frac{1}{3-k}\times\frac{3I+2I^2}{1+2I+I^2} \\ d_1 &= \frac{1}{2}\times\frac{3I+2I^2}{1+2I+I^2} \\ d_2 &= 1\times\frac{3I+2I^2}{1+2I+I^2} = 2d_1 \quad . \end{aligned} \tag{2.58}$$

Similarly we can analyze $n = 3$ to :

$$\begin{aligned} d_k &= \frac{(3I-1)\times(1+I)^3+1}{I(4-k)\times(1+I)^3} \\ &= \frac{1}{4-k}\times\frac{I(6+8I+3I^2)}{1+3I+3I^2+I^3} \\ d_1 &= \frac{1}{3}\times\frac{I(6+8I+3I^2)}{1+3I+3I^2+I^3} \\ d_2 &= \frac{1}{2}\times\frac{I(6+8I+3I^2)}{1+3I+3I^2+I^3} = \frac{3}{2}d_1 \\ d_3 &= 1\times\frac{I(6+8I+3I^2)}{1+3I+3I^2+I^3} = 2d_2 = 3d_1 \quad . \end{aligned} \tag{2.59}$$

Interestingly enough, it is easy to show that

$$\frac{I}{I+1} < \frac{3I+2I^2}{1+2I+I^2} = \frac{3I+2I^2}{(1+I)^2} < \frac{I(6+8I+3I^2)}{1+3I+3I^2+I^3} = \frac{I(6+8I+3I^2)}{(1+I)^3} \tag{2.60}$$

what to make of this?

Look at the dependence of d_k with d_{k-1} etc. to determine defaults at each level. Then use for I

Limits & Approximations:

Equation 2.55 can be assessed for $I \to 0$ using the Taylor expansion 2.10

$$\begin{aligned} \lim_{I\to 0} d_k &= \lim_{I\to 0}\frac{(nI-1)(1+nI+\dots)+1}{I(n+1-k)(1+nI+\dots)} \\ &= \frac{n^2I^2}{I(n+1-k)(1+I)} \\ &= \frac{n}{n+1-k}\times\frac{nI}{1+nI} \quad , \end{aligned} \tag{2.61}$$

which we can then use to determine I as a function of d_k for very small I:

$$d_k \times (1+nI) = \frac{n}{n+1-k}nI \tag{2.62}$$

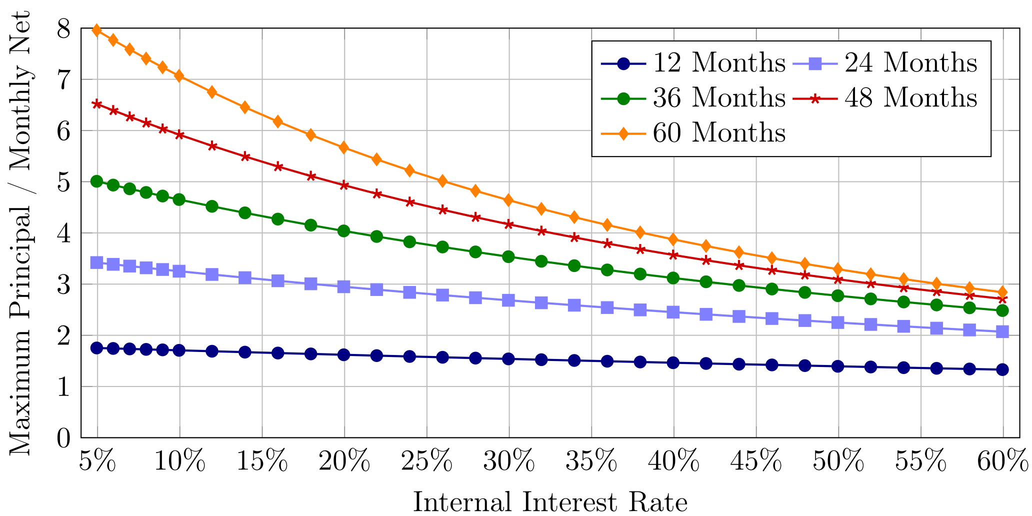


Figure 2.23: Maximum borrow amount as multiple of monthly net income for internal interest rates from 5% through 60% for various terms given for a payment to income ratio of $\iota = 15\%$. It scales linearly with ι.

and therefore

$$I = \frac{d_k}{n\left[\frac{n}{n+1-k} - d_k\right]} \quad . \tag{2.63}$$

Keep in mind though, that we would want to stay far far below the break even default rate calculated in order to run a profitable business. Figuring out how far below would have to be determined by the business goals.

work out math with marketing cost, reporting cost, ops cost, default rates and prepayments, ignoring recoveries, cost of money

Example(s):

For the loan given in table 2.2 We would need 1.66 regular loans to be paid off for each loan that defaults initially and 1.33 for each loan that defaults after three installments just to break even on the product without making any money. There goes the challenge! **How did I do that?**

compute ideal APRs from known d, n, k (explicit expression?), compare to model score, results and business rules

The above considerations show that break even point default rate, number of profitable payment and number of loans required to make up for defaulting ones **do not** depend on the principal loan amount. However, the amount of money we make **does**.

Let us venture into a few considerations on the principal loan amount. It seems to be a good idea to relate that to a few business rules regarding minimum and maximum monthly payments, minimum and maximum loan amounts in relation to the monthly net available income, the amount paid for acquiring the customer, regulatory requirements imposed by state, country, third parties etc. This will end up in a set of competing business rules that need analysis.

put this in a table / graphical representation

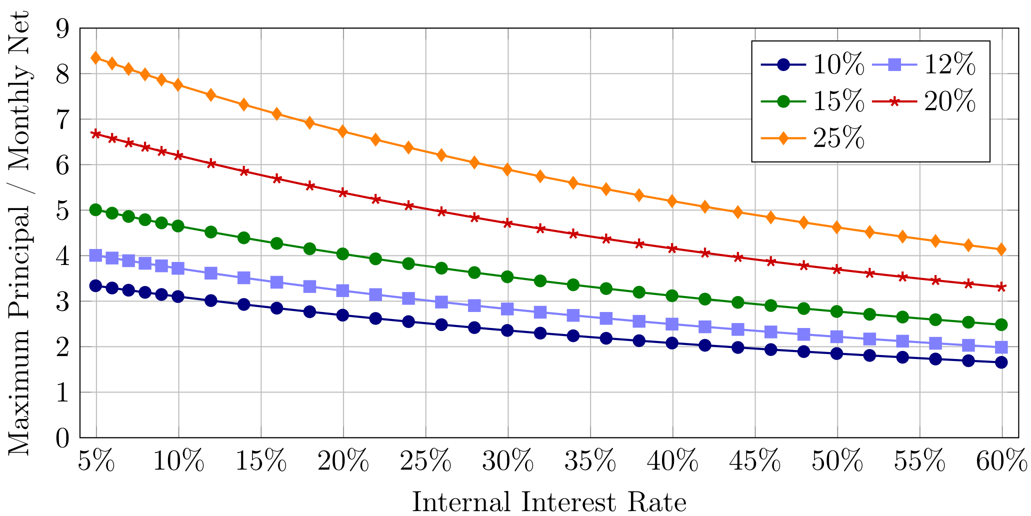


Figure 2.24: Maximum borrow amount as multiple of monthly net income for internal interest rates from 5% through 60% for a 36 months terms given for various payment to income ratios.

2.3.3 Designing the Loan Product

willingness to pay vs. ability to pay, Expand bla bla blub

So, how much principal P can we give give, if the installment amount α equals a fixed percentage ι of the monthly net income M. We can easily determine that from equation 2.6.

$$\alpha = \iota M \quad \rightarrow \quad \frac{P}{M} = \iota \frac{1-(1+I)^{-n}}{I} \tag{2.64}$$

which, of course, assumes we are not restricted by anything else such as hard caps, hard minima, minimum monthly payment amounts or other income independent parameters as well as arbitrarily set values of loan to income. All of which will be present in the actual business implementation.

do math for maximum P for doubling term or adding 12 months - then how much is the payment going down and the overall going up

Example(s):

To get a feeling for this relation, let us consider how much of principal you could get if you want to determine a 30 year mortgage at 3.5% nominal interest rate (assuming no additional fees) and not commit any more than 43% of your monthly net income. That gives:

$$P/M = 0.43/0.035 \times 12 \left[1-(1+0.035/12)^{-360}\right] \approx 96, \tag{2.65}$$

or ≈ 8 times your yearly net income. Which would drop to ≈ 5.5 if your interest rate doubled over the same term and ≈ 4 if you do shoot for a double rate half term solution.

Given a hard cap P_h on principal, how much monthly income M_h would be required to hit that for any given ι? Easily calculated:

$$M_h = \frac{P_h}{\iota} \frac{I}{1-(1+I)^{-n}} \quad . \tag{2.66}$$

graphics and table, very interesting behavior, maybe on example of live rules in UK

The minimum loan amounts P_m as well as the minimum payment amounts α_m, ι_m also cause requirements for minimum monthly income M_m and minimum loan amount respectively.

$$\begin{aligned} P_m \geq \alpha_m \frac{1-(1+I)^{-n}}{I} \quad &; \quad P_m \geq \frac{\iota M}{I}\left[1-(1+I)^{-n}\right] \\ M_m \geq \alpha_m \quad &; \quad M_m \geq \frac{P_m}{\iota_m}\frac{I}{1-(1+I)^{-n}} \end{aligned} \tag{2.67}$$

write this up

2.4 Loan Life Cycle

Up until this point we have dealt with loans that consist of n installments of equal length. In practice, that may not be the case, actually it almost never is because of the difference between months / installments in number of days in them, the bank holidays involved an / or even the effects of a leap year. So how does the loan get determined?

2.4.1 Terminology

- An **installment** is a sum of money due as one of several equal payments for something, spread over an agreed period of time. For us this is the point of reference of what should happen that the customer agreed to by signing the contract. Comparing payments form customer to the original installment amount we can figure out over / under payment and using the installment date we determine whether or not the payment got to us in time.
- **Payments** are the materialized representation of how much the customer has paid us (or will pay us) on a specific date. In real life, the customer's payments do not need to (and often don't) match what the payment-schedule-as-defined-by-the-installments says. Payments that the customer is scheduled for and that we attempt to take may be returned up to 10 days later. A customer may also call in and schedule additional payments. Payments are a record of these.
- **LoanTasks** bridge a gap between installments and payments. Through LoanTask status (created, completed, cancelled, returned), Loan Tasks reflect as-much-as-possible current up-to-the-minute reality.
- When a loan task moves money between "accounts" (also commonly referred to as "buckets"), it does so through a set of **Payment Transactions**. PaymentTransactions belong to LoanTasks. Each PaymentTransaction is an atomic movement of an amount of money between two buckets: an account to 'credit' and an account to 'debit'.
- Each Product has what is termed "**the waterfall**", which is the ordering of accounting buckets to which payment amounts are applied.

  See `PaymentTransaction.waterfall`.

bla bla

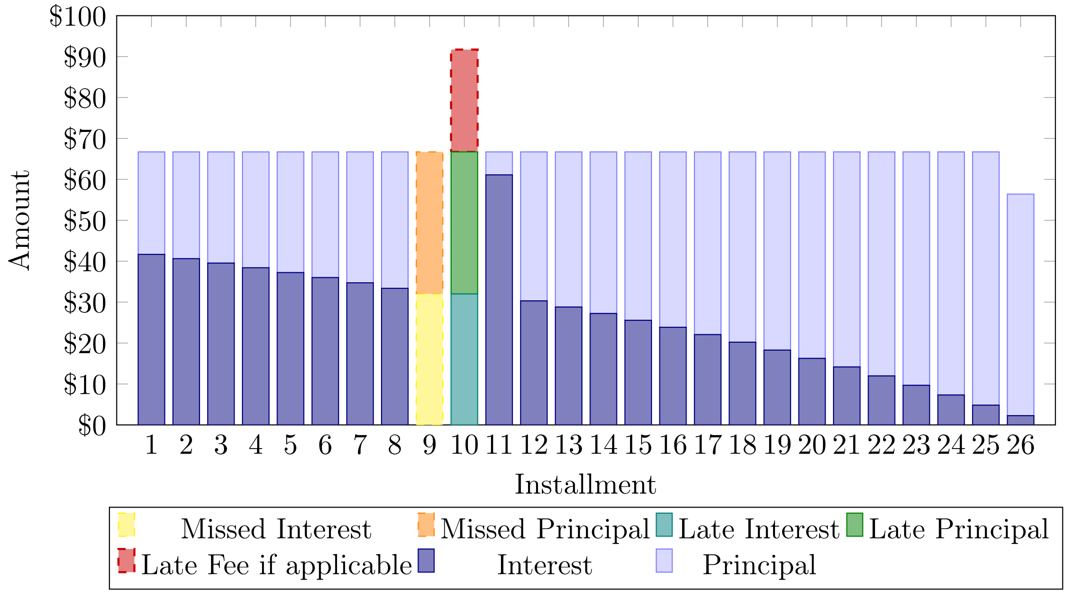


Figure 2.25: Shares of various additional account buckets, interest and principal for the loan discussed above, 24 months, $1000, 50% APR. We assume the customer misses payment 9 as either a courtesy payment or with a late fee, in which case we would try to collect the fee with installment 10.

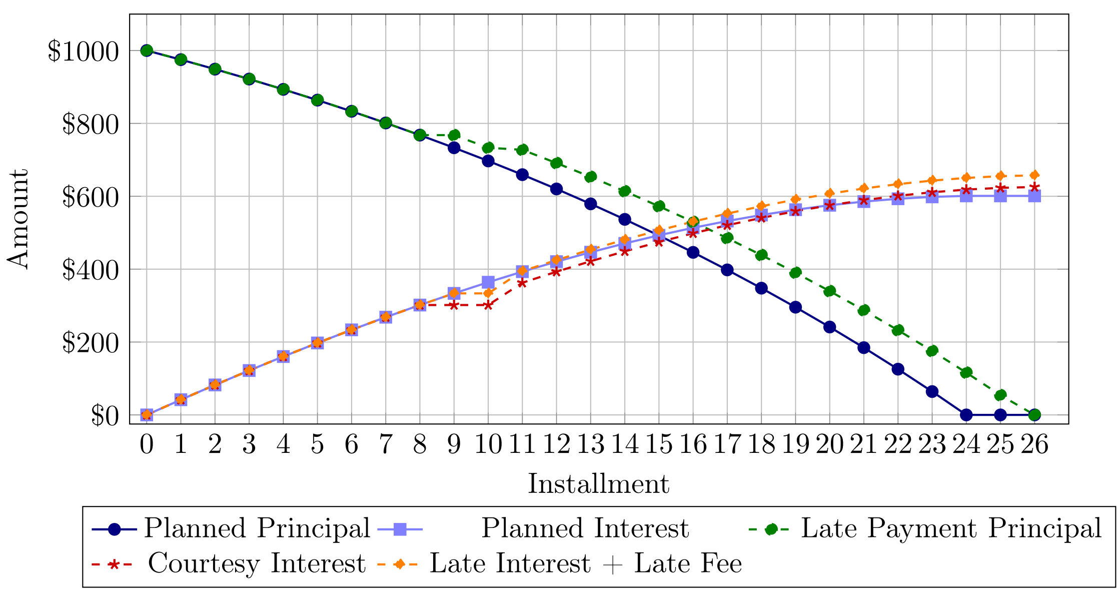
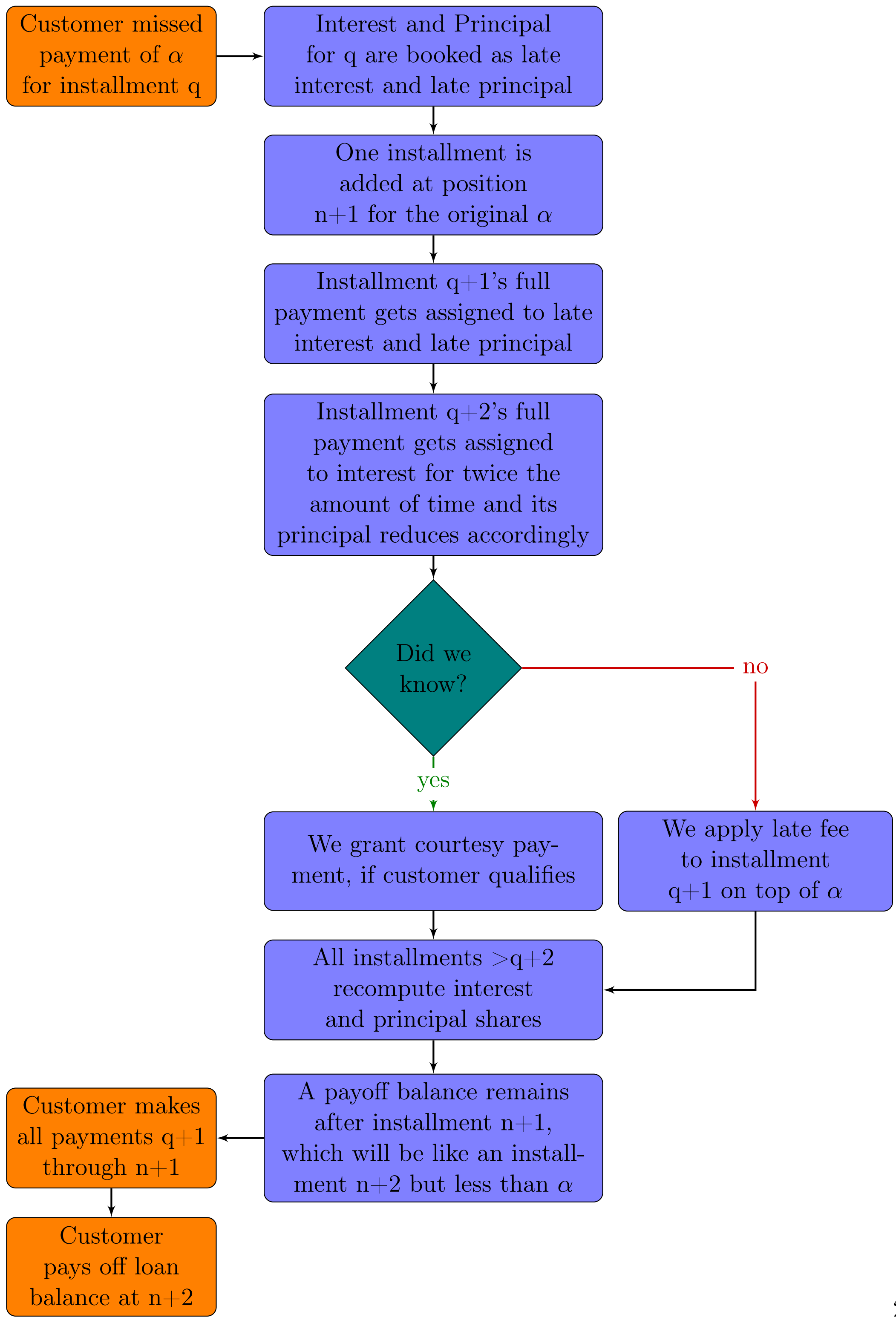


Figure 2.26: Principal and Interest for the case of a missed payment number 9 out of 24 for a 50%, $1,000 loan. Noteworthy, the amount of interest collected by installment 24, the original term, already exceeds the amount we would have made.



2

2.5 References

Wikipedia: Loans

Updated: 2018/01/05 13:11

- Wikipedia: Installment Credit

Chapter 3

Credit Cards

Credit cards as known to the public are a small plastic card that provides consumers with a method of payment in the form of a revolving credit line. It is by far the most common form of consumer credit by count as seen in figure 1.3 with some 450M accounts in the US alone but also the one with the lowest average balance (≈$1,800) - see figure 1.4 - and overall balances of ≈$800B, see figure 1.2 at an overall limit in the order of $3T.

Table 3.1 summarizes the quarterly issuance statistics as observed in the TransUnion Industry Insights Report of Q2 2017 for all issuance in Q1 2017. Monthly originations and average credit lines are given in figure 3.1 provide by the Consumer Financial Protection Bureau. The same source provides time dependent monthly issuance data by credit quality as given in figures 3.2 and 3.3. While the definitions of credit quality levels differ quite a bit between TransUnion and the CFPB, it is noteworthy that the numbers of newly created accounts are quite similar between the levels of credit quality, but the limits increase dramatically with increasing quality of credit.

Typically, cards contain a form of identification, a photo or signature that authorizes the person named on it to charge for goods and services for which the cardholder is billed periodically.

Examining the components that make a credit card, we start with *revolving credit*. In contrast to installment loans, **revolving credit** does not have a fixed term or fixed monthly payments. Instead, payment amounts will depend on the balance drawn (used). Credit is granted up to a fixed, pre-set amount called the limit and can be withdrawn, repaid and redrawn in any manner and any number of times until the arrangement expires

Table 3.1: Credit Card issuance volume by credit quality in Q1 2017 according to the TransUnion Industry Insights Report Q2 2017. Balances are end-of-quarter values for cards issued in the quarter.

Credit Quality Level	Total	Subprime	Near Prime	Prime	Prime Plus	SuperPrime
Vantage Score 3 Range		300-600	601-660	661-720	721-780	781-850
Accounts (M)	15.2	2.7	2.9	3.3	3.2	3.1
Balance ($B)	18.0	2.3	4.0	4.7	4.6	2.4
Limit ($B)	82.9	2.4	6.9	13.1	24.5	36.0
Average Balance ($)	1,700	900	1,700	1,000	2,300	1,500
Average Limit ($)	5,800	1,000	2,600	4,200	8,200	12,200

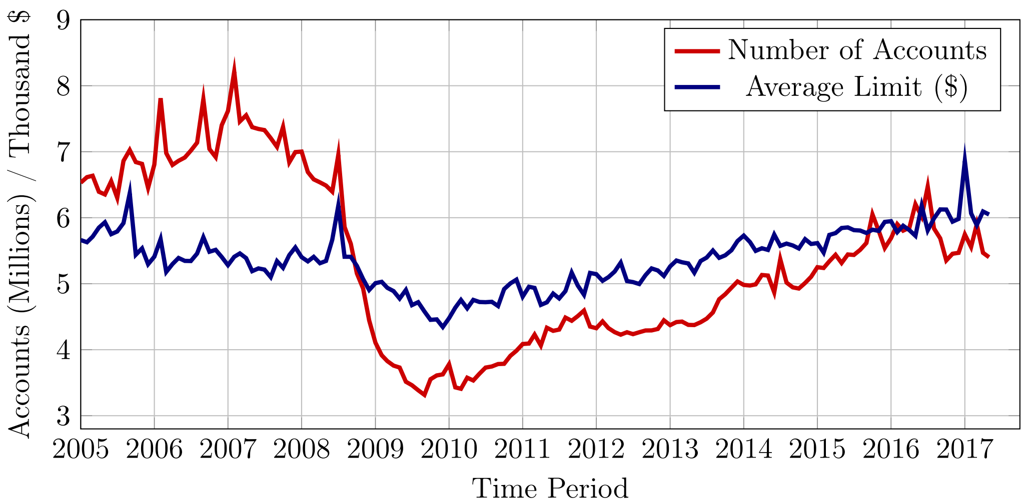


Figure 3.1: Number and average limit of credit cards issued by month according to CFPB.

as the borrower sees fit. Common forms of revolving credit are credit cards, any lines of credit such as home equity lines of credit, overdrafts but also some debt facilities for enterprises.

Typical characteristics of revolving credit are:

- The borrower may use or withdraw funds up to a pre-approved credit limit.
- The amount of available credit decreases and increases as funds are borrowed and then repaid.
- The credit may be used repeatedly.
- The borrower makes payments based only on the amount he or she has actually used or withdrawn, plus interest.
- The borrower may repay over time (subject to any minimum payment requirement), or in full at any time.
- In some cases, the borrower is required to pay a fee to the lender for any money that is undrawn; this is especially true of corporate bank revolving-credit loans.

A revolving loan is a particularly flexible financing tool as it may be drawn by a borrower by way of straightforward loans, but it is also possible to incorporate different types of financial accommodation within it - for example, it is possible to incorporate a letter of credit, swingline or overdraft within the terms of a revolving credit loan. This is often achieved by creating a sublimit within the overall loan, allowing a certain amount of the lenders' commitment to be drawn in the form of these different facilities.

Unlike installment loans revolving credit lines can be subject to changes during their life cycle. In particular, variable interest rates are common that follow the prime rate plus a certain spread (see figure 1.11). Additionally, account management policies such as line increases and decreases become relevant.

The other part that makes a credit card is its function as a form of payment. In the US credit cards are almost universally accepted as a form of payment, in other highly developed countries too. In less developed countries or countries with stricter

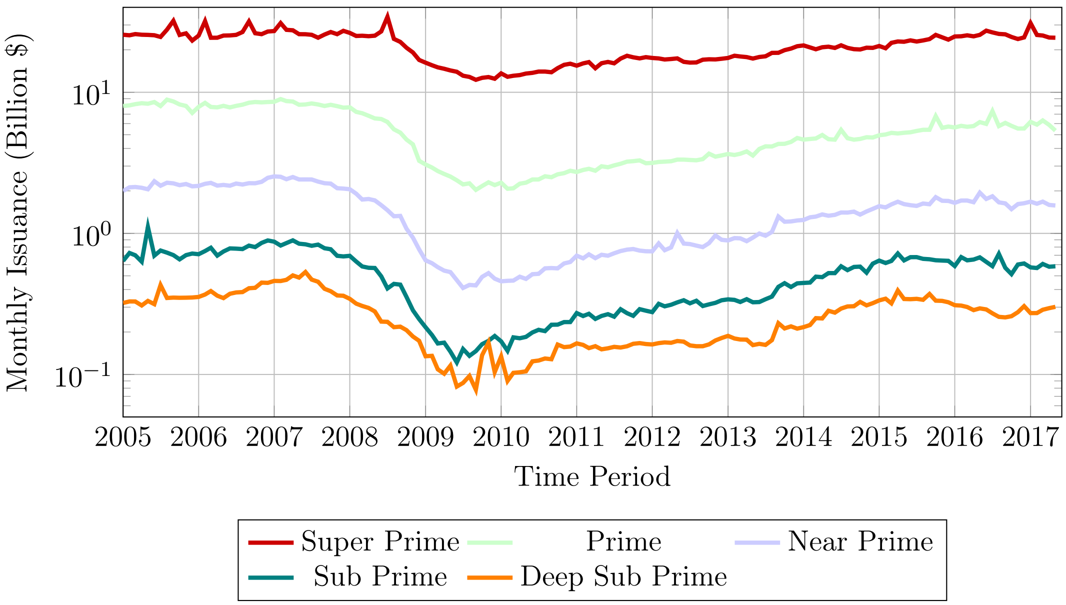


Figure 3.2: Credit card issuance volume by month ($ limit) according to CFPB. Definitions of prime levels: Deep subprime <580, subprime 580-619, near prime 620-659, prime 660-719, superprime 720$^+$.

bankruptcy laws such as Germany, credit cards are much less used. In the payment system the issuer pays the merchant when the borrower used the credit card for payment and will subsequently bill the money to the borrower as part of the monthly billing statement which the borrower then can pay in full in which case no interest charge accrues or partially in which case interest will be charged. In addition to costs covered by the borrower, the merchant will pay a fee for using the payment method to the issuer / network.

As credit cards have become ubiquitous, the relative freedom of spending lead to consumers outspending their earnings which quickly resulted in fees and defaults. Credit card products are therefore one of the most highly regulated financial products. Additionally, responsible use of revolving credit, especially credit cards, is one of the strongest indicators / predictors of future payment behavior and features prominently in almost every single credit score in existence, see chapter 4.

To determine the details of a credit card product the customer is interested in the following:

- Questions on limit
- What is the **amount** of minimum payment?
- Questions on cash, overlimit etc.?
- Questions?

As **AvANT** we would like to know a few more numbers to develop a profitable business **Modify more of this**

- **How much** money do we make?
- **When** do we start making money (after which payment) for each credit card in the best of cases (happy path)?

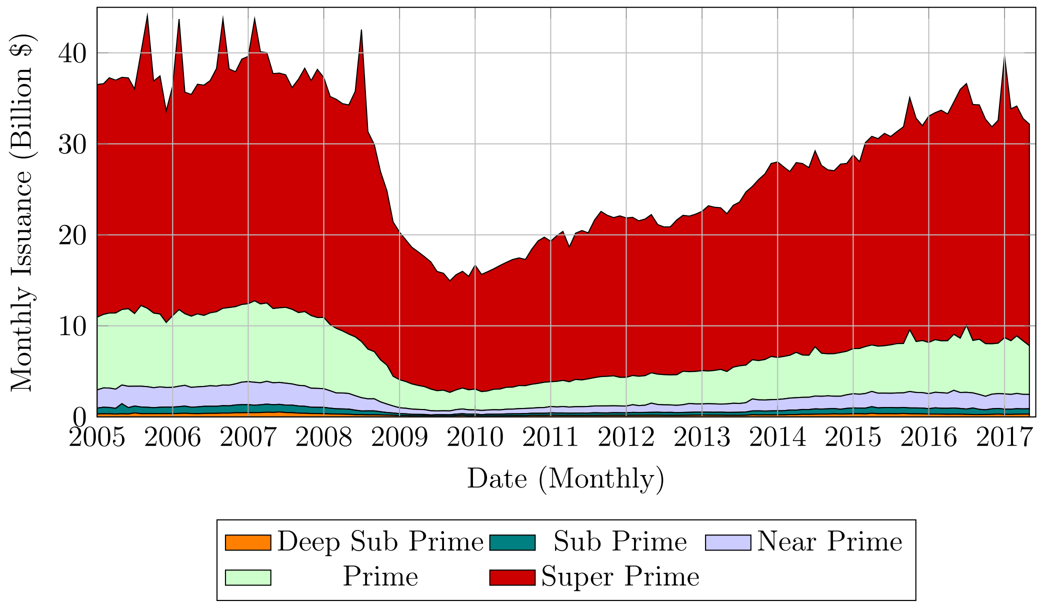


Figure 3.3: Credit card issuance in a linear stacked plot. The data is the same as in figure 3.2. The dominance by balance volume of super-prime is clearly noticeable. By count however, all segments have appreciable numbers behind them.

- **What** can we do in account management to incur maximum profits per card?
- **How many** cards of the same type need to be paid off fully in order to maintain a profitable product given a delinquency on any given payment?
- more questions

3.1 History and Market Overview

Credit, which was once the sign that a person had trouble meeting his bills, has taken on a glamorous new meaning in recent years. Now a man with a credit card can rent a plane or boat or car, live it up in nightclubs, take a safari to Africa and even get a Kelly Girl for temporary office help. Why? Because of the Credit Card Game.

Time magazine 1958

1850 American Express is founded as an express delivery company

1865 Department Stores issue first Charge Coins

1882 American Express introduces Money Orders

1891 American Express offers traveller's cheques

1899 first ID theft on a charge account

1914 Department store credit becomes ubiquitious

1920s Individual firms (oil, hotel industries) issue credit cards to consumers for purchases at company outlets

1935 Charge Plates at peak popularity through 1950

1946 Charg-it launched by Biggins Bank

1950 Diners Club established first universal credit card

1952 Franklin National Bank introduces first credit card with *buy now, pay later, with interest*

1951 First Patent with phrase *credit card*

1958 American Express establishes a *travel and entertainment card*

1958 Hilton Hotels expand *Carte Blanche* from pure affinity to credit card

1958 BankAmericard introduces the bank credit card system in California, in which the bank credits a merchant and will collect the money later from the consumer. Further states follow from 1966

1959 American Express introduces plastic charge cards, others follow in the early sixties

1966 Bankamericard licensed as first general purpose card

1966 Interbank Card Association founded

1969 International Business Machines (IBM) develops magnetic stripes for plastic cards

1970 Fair Credit Reporting Act

1974 Fair Lending Billing Act, Equal Credit Opportunity Act

1976 BankAmericard rebranded to VISA

1977 Fair Debt Collection Practices Act

1979 Interbank Card Association rebrands to MasterCard

1980 VISA and MasterCard move to magnetic stripe technology

1980 Debit cards introduced

1981 American Airlines unveils first rewards program

1983 American Express introduces first affinity credit card

1985 First cash back product introduced as Discover by Sears

1986 First EMV enabled cards introduced in France

1990 AT&T issues first no-annual-fee-for-life card

2002 Mini Cards

2007 First contactless cards issued by Barclaycard in the UK

2007 Personalized credit cards

2008 Discover acquires Diners Club

2009 Credit CARD Act of 2009 establishes extensive consumer protection and expands the *Consumer bill of rights*

2015 EMV use in US cards becomes dominant

Top issuers, top networks stats, etc.

Table 3.2: US Merchant Card Fees 2016 by network according to the Nilson Report

Network Type	Fee Revenue ($B)	Purchase Volume ($B)	W. avg (%)
Visa/Mastercard Credit	47.53	2,242.02	2.12
American Express	16.41	695.44	2.36
Visa/Mastercard Debit	15.21	2,083.65	0.73
PIN Debit	3.39	505.73	0.67
Private Label Credit	3.37	287.86	1.17
Discover	2.48	121.47	2.04
Total	88.39	5,936.17	1.49

Table 3.3: Volume of transactions for global general purpose cards by network according to the Nilson Report. pp indicates prepaid cards, Merch a merchandise transaction/ Purchase

	Volume ($B)			Transactions (Bs)			Cards
Brand card type	Total	Merch	Cash	Total	Merch	Cash	(Ms)
VISA Credit	4,074.19	3,778.99	295.20	50.82	49.46	1.36	1,007.6
Union Pay credit	3,235.35	3,189.97	45.39	19.73	19.52	0.21	465
MasterCard credit	2,534.58	2,349.43	185.15	33.87	33.09	0.78	781
American Express credit	1,037.50	1,023.76	13.74	7.35	7.23	0.12	109.9
JCB credit	241.22	234.01	7.21	3.02	2.96	0.06	103.2
Discover/ Diners club	160.98	148.73	12.24	2.41	2.34	0.07	56.8
Credit card totals	11,283.82	10,724.89	558.93	117.2	114.6	2.6	2,523.5
Union Pay debit & pp	6,415.81	5,312.70	1,103.11	22.71	18.81	3.9	5,660
VISA debit & pp	6,051.28	3,403.02	2,648.27	110.77	89.56	21.21	2,079.3
MasterCard debit & pp	2,292.85	1,165.01	1,127.84	44.43	34.2	10.23	887.9
Debit card totals	14,759.94	9,880.73	4,879.22	177.91	142.57	35.34	8,627.2
VISA totals	10,125.47	7,182.01	2,943.47	161.59	139.02	22.57	3,086.9
MasterCard totals	4,827.43	3,514.44	1312.99	78.3	67.29	11.01	1,668.9
VISA +MasterCard totals	14,952.90	10,696.45	4,256.46	239.89	206.31	33.58	4,755.8
Union Pay totals	9,651.16	8,502.67	1,148.50	42.44	38.33	4.11	6,125
Grand totals	26,043.76	20,605.62	5,438.15	295.11	257.17	37.94	11,150.7

Table 3.4: Per card statistics for global general purpose cards by network according to the Nilson Report. d & pp indicates debit and prepaid cards, Merch is for merchandise transactions/ purchases. The data is an elaboration of table 3.3 and follows the same description.

	Transactions			**Avg. Transaction ($)**			**Volume ($)**		
Brand card type	**Total**	**Merch**	**Cash**	**Total**	**Merch**	**Cash**	**Total**	**Merch**	**Cash**
VISA Credit	50.44	49.09	1.35	80.17	74.36	5.81	4,043.46	3,750.49	292.97
Union Pay credit	42.43	41.98	0.45	163.98	161.68	2.30	6,957.74	6,860.15	97.61
MasterCard credit	43.37	42.37	1.00	74.83	69.37	5.47	3,245.30	3,008.23	237.07
American Express	66.88	65.79	1.09	141.16	139.29	1.87	9,440.40	9,315.38	125.02
JCB credit	29.26	28.68	0.58	79.87	77.49	2.39	2,337.40	2,267.54	69.86
Discover/ Diners	42.43	41.20	1.23	66.80	61.71	5.08	2,834.15	2,618.49	215.49
Credit card totals	46.44	45.41	1.03	96.28	91.51	4.77	4,471.50	4,250.01	221.49
Union Pay d& pp	4.01	3.32	0.69	282.51	233.94	48.57	1,133.54	938.64	194.90
VISA d& pp	53.27	43.07	10.20	54.63	30.72	23.91	2,910.25	1,636.62	1,273.64
MasterCard d & pp	50.04	38.52	11.52	51.61	26.22	25.38	2,582.33	1,312.10	1,270.23
Debit card totals	20.62	16.53	4.10	82.96	55.54	27.43	1,710.86	1,145.30	565.56
VISA	52.35	45.04	7.31	62.66	44.45	18.22	3,280.14	2,326.61	953.54
MasterCard	46.92	40.32	6.60	61.65	44.88	16.77	2,892.58	2,105.84	786.74
VISA +MasterCard	50.44	43.38	7.06	62.33	44.59	17.74	3,144.14	2,249.14	895.00
Union Pay	6.93	6.26	0.67	227.41	200.35	27.06	1,575.70	1,388.19	187.51
Grand totals	26.47	23.06	3.40	88.25	69.82	18.43	2,335.62	1,847.92	487.70

Table 3.5: Top Card Issuers in the US according to Valuepenguin.com

	Totals in $B/Ms			Avg. per consumer in $/%		
Issuer	Balance	Consumers	Limit	Balance	Limit	Utilization
Citibank	146.22	95.4	815	1,532.70	8,542.98	18
JP Morgan Chase	131.23	82.8	685	1,584.90	8,272.95	19
Capital One	98.41	62.1	411	1,584.70	6,618.36	24
Bank of America	92.27	58.2	457	1,585.40	7,852.23	20
Discover	61.37	38.7	254	1,585.79	6,563.31	24
Synchrony	58.13	36.7	405	1,583.92	11,035.42	14
American Express	40.10	62.7	519	639.55	8277.51	8
Wells Fargo	36.70	23.2	153	1,581.90	6,594.83	24
Barclays	25.93	16.3		1,590.80		
US Bancorp	21.75	13.7	148	1,587.59	10,802.92	15
World Financial			116			

3.1.1 Types of Plastic (Credit) Cards

- **Bankcards** (VISA, MasterCard, American Express, Discover, China Union Pay, JCB)
    - **Affinity cards:** Offered in conjunction between a card issuer and a non-financial institution with which consumers have an affinity such as universities, sport franchises or nonprofit organizations. Often feature special discounts or deals for using their cards as opposed to non-affinity cards. Revenue is typically shared between the card issuer and the non-financial institution.
    - **Secured cards:** Provide customers with bad credit history or no history with a way to (re)build credit history by offering a plastic credit card with a pre-set limit equivalent to the deposited amount the customer secured the card on. The payment behavior on secured cards is substantially better than on unsecured cards, especially in the lower credit score range.
- **Charge cards (Travel & Entertainment Cards):** Require full payment of charges at the end of the payment cycle by the statement due date. This is not a revolving credit product (no carried balances) and does not charge an interest rate.
- **Co-Branded Cards:** Sponsored by two parties, typically a retailer (e.g. Airline, Department Store, Gas Retailer, etc.) and a bank or network. Typically features some sort of merchandise discounts or rewards program. Revenue is usually shared between the two parties on much more even footing than with affinity cards.
- **Private Label Cards:**
    - Closed loop without use of any interchange network,
    - Can only be used at specific retailer and nowhere else,
    - Lenders typically offer promotions on larger purchases through this type of card: Deferred payment (No Payment for 12 months), Deferred Interest (No Interest if paid in full in 12 months) or Purchase discounts for using the card.
- **Debit Cards:** though technically not a credit card it is a ubiquitous type of plastic card
    - Provides access to customers current deposit account sometimes with an overdraft feature
    - Two types of transactions: PIN based online access to deposit account and offline access through a network (MasterCard, VISA)
    - Revenue for the issuer comes from fees on PIN based transactions and interchange percentage of network transactions
    - Credit losses are non-existent, frauds losses smaller than 50 bps

3.1.2 Common Features of Credit Cards

Revenue Flow / Involved Parties

lots of text, some example, nice shiny flow chart

Authorization / Settlement flow

lots of text, some example , nice shiny flow chart

Balances

Mostly three types of balances:

- Merchandise
- Balance transfer
- Cash

All of these balances can and usually will be treated differently. It is also possible and common to subject either of these balances to promotional arrangements.

Statements

3.2 Consumer Considerations

Explain terms: Limit / Line, Balance, Utilization, ...

3.2.1 How is interest assessed?

There are two forms of interest that can be charged: Fixed Rate Interest and Variable Rate Interest. In fixed rate interest the rate is constant and does not change during the life of the card unless there is a change in terms. As of 2009, card rates advertised as fixed must remain fixed for a period of at least 12 months and can only be with 45 days prior notice. Variable rate interest is tied to another specific rate called an index, e.g. the Prime Rate as set by the Wall Street Journal or the Federal Funds rate. If the index rate rises by 25 bps, the interest rate charge to the balance of the account will rise by 25 bps. Variable Interest Rates are pre-dominant in the market. Variable interest rates are frequently disclosed as

$$\text{Interest Rate} = \text{Prime} + \text{Spread},$$

so if for example the prime rate sits at 4.25% and the spread disclosed in the cardholder agreement is 17.74%, the resulting APR is 21.99%. An illustration of the historic Prime rate, Federal Funds rate and averaged credit card rates is given in figure 1.11.

3.2.2 Billing cycle

The **billing cycle** is is a specific, recurring time period between billing statements. Under the CARD Act of 2009, due dates must be the same day every month, and payments that come due dates that fall on weekends or holidays are not subject to late fees.

A **billing statement** or monthly statement is a written record prepared by a financial institution, usually once a month, listing all credit card transactions for an account,

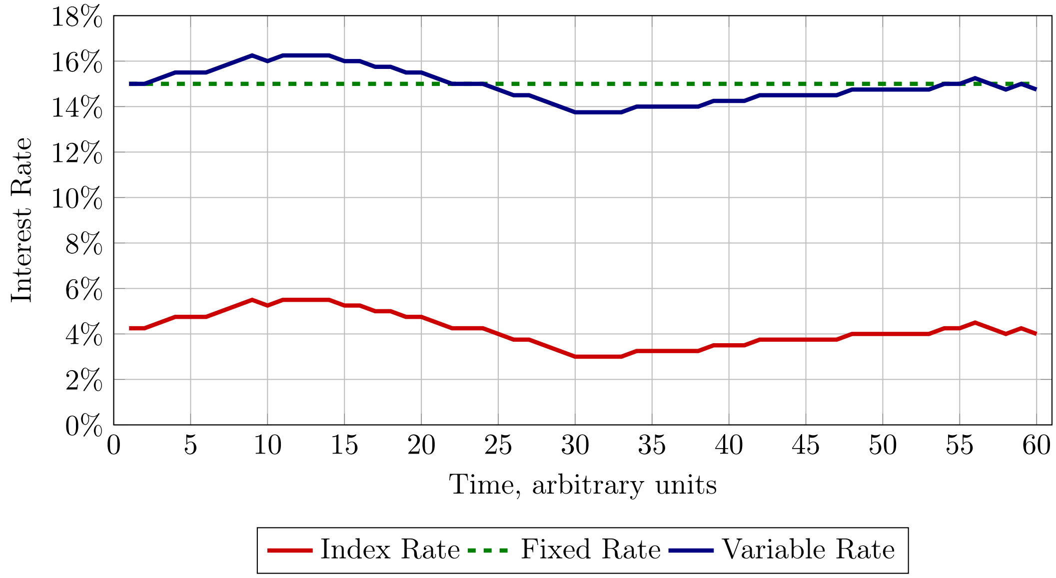


Figure 3.4: Index, Fixed and Variable Rates, both set to 15% at the first cycle. This illustrates the difference between fixed and variable rates over a period of time where the index rate fluctuates but no changes in terms have been made.

including purchases, payments, fees and finance charges. It will be provided by mail or electronically, depending on customer choice.

The last day of a billing cycle is called **statement cycle date** or cycle date. On this day, all balances, finance charges and fees for the ending cycle will be assessed and a statement will be produced.

A very important day in a billing cycle is the **payment due date** by which a payment has to be made by the customer that covers at least the minimum payment due amount such as to keep the account current and avoid delinquent status and late fees. As of 2009 the due date must be at least 21 days after the statement cycle date.

tons of examples, statement cycle date, min pay due date

3.2.3 Average Daily Balance method

The Average Daily Balance (ADB) method is by far the most common method of computing interest in the credit card space of today. It is determined by adding the balance at the end of each day in a billing cycle and divide it by the number of days in a billing cycle.

$$\text{Average Daily Balance} = \frac{\sum_{\text{Days in Cycle}} (\text{Balance at End of Day})}{\text{Days in Cycle}} \tag{3.1}$$

The interest in a given cycle works out to

$$\text{Cycle Interest} = \text{Average Daily Balance} \times \frac{\text{Annual Percentage Rate}}{\text{Number of Cycles in a Year}} \quad . \tag{3.2}$$

A legally more precise application is

$$\text{Cycle Interest} = \frac{\text{ADB} \times \text{APR} \times \text{Days in Cycle}}{\text{Days in a Year}} \quad . \tag{3.3}$$

It applies to all three types of balances individually. If a customer account carries a negative balance (the customer paid more than the balance), the daily balance for all days with negative balance is set to zero, it does not counter positive balances.

Example(s):

We shall assume that the billing cycle has 30 days, set the minimum payment due date to the 24th day of the cycle and look at the end of day balances each day.

Single purchase on a zero balance:

Day	Activity	Cash Flow	Balance
0	Starting Balance		$0.00
15	Purchase	**$100.00**	**$100.00**
24	Minimum Payment	**-$25.00**	**$75.00**
30	Statement Cycle Date		**$75.00**

So we have 15 days with $0 balance, followed by 9 days of $100, 6 days of $75 to result in the end of billing cycle balance of $75. In this case the ADB works out to:

$$\begin{aligned} \text{ADB} &= (15 \times \$0 + 9 \times \$100 + 6 \times \$75)/30 \\ &= \$1,350/30 \\ &= \$45 \end{aligned}$$

Multiple purchases and multiple payments with starting balance:

Day	Activity	Cash Flow	Balance
0	Starting Balance		$80.00
5	Purchase	**$100.00**	**$180.00**
	Payment	**-$150.00**	**$30.00**
8	Purchase	**$20.00**	**$50.00**
24	Payment	**-$25.00**	**$25.00**
30	Statement Cycle Date		**$25.00**

Importantly, on day 5 both a purchase and a payment occurred so the end-of-day balance is the sum of the balance the day prior, plus purchase less payment.

$$\begin{aligned} \text{ADB} &= (5 \times \$80 + 3 \times \$30 + 16 \times \$50 + 6 \times \$25)/30 \\ &= \$1,440/30 \\ &= \$48 \end{aligned}$$

If the customer had paid both payment amounts combined on day 24 his ADB would have resulted in

$$\begin{aligned} \text{ADB} &= (5 \times \$80 + 3 \times \$180 + 16 \times \$200 + 6 \times \$25)/30 \\ &= \$4,290/30 \\ &= \$143 \end{aligned}$$

or almost 3x the ADB he had with the earlier payment. As can be seen in this example, the day in the cycle the payment is made does significantly influence the average daily balance on which interest gets assessed. Earlier is better even if it is not possible to pay down the entire balance or even the minimum payment. Partial payments earlier in the cycle do modify the interest cost significantly, especially when the payment amount is close to a full payoff. If for example, the starting balance had been $580 and all cash flows the same, the average daily balance would turn out to be $548 for the early payment and $643 for paying at the due date. That is no longer a 3x difference but still appreciable.

Multiple purchases, cash advances, returns and over-payments in the cycle:

			Balances		
Day	**Activity**	**Cash Flow**	**Merch.**	**Cash**	**Fees**
0	Starting Balance		$50.00	$0.00	$0.00
5	Purchase	$100.00	$150.00	$0.00	$0.00
10	Payment	-$200.00	-$50.00	$0.00	$0.00
14	Purchase	$70.00	$20.00	$0.00	$0.00
16	Cash Advance	$50.00	$20.00	$50.00	
	Cash Advance Fee	$5.00			$5.00
20	Purchase	$50.00	$70.00		
	Cash Advance	$50.00		$100.00	
	Cash Advance Fee	$5.00			$10.00
21	Return Purchase	-$20.00	$50.00	$100.00	$10.00
24	Payment	-$160.00	$0.00	$0.00	$0.00
27	Purchase	$50.00	$50.00	$0.00	$0.00
30	Statement Cycle Date		$50.00	$0.00	$0.00

In this case we do have two interest bearing balances, cash and merchandise / purchase, as well as a fee component that will be reflected in the statement. The two balances are not necessarily and commonly not subject to the same APR, so the ADB has to be computed separately for both cases.

Cash balance calculation:

$$\begin{aligned} \text{ADB}|_{\text{Cash}} &= (16 \times \$0 + 4 \times \$50 + 4 \times \$100)/30 \\ &= \$600/30 \\ &= \$20 \end{aligned}$$

Merchandise Balance Calculation:

This calculation must also consider the overpayment early in the cycle. Negative balances are regarded as zero in virtually all cases.

$$\begin{aligned}
\text{ADB}|_{\text{Merchandise}} &= (5\times\$50 + 5\times\$150 + 4\times\$0 + 6\times\$20)/30 \\
&\quad +(1\times\$70 + 3\times\$50 + 3\times\$0 + 3\times\$50)/30 \\
&= \$1,490/30 \\
&= \$49.67
\end{aligned}$$

Had the balance be allowed to be negative, the ADB would be \$2.67 less.

Side Note:

The average daily balance method is mathematically equivalent to applying simple daily interest to the balance at the end of each day. Let B_i be the balance at the end of day i in the cycle. The simple interest calculation for the interest in the period would result in

$$\begin{aligned}
\text{Cycle Interest} &= \sum_i \left(B_i \times \frac{\text{Annual Percentage Rate}}{\text{Number of Days in a year}} \right) \\
&= \sum_i \left(B_i \times \frac{\text{Annual Percentage Rate}}{\text{Cycles per year} \times \text{Days per cycle}} \right) \\
&= \sum_i \left(\frac{B_i}{\text{Days per cycle}} \times \frac{\text{Annual Percentage Rate}}{\text{Cycles per year}} \right) \\
&= \left(\frac{\sum_i B_i}{\text{Days per cycle}} \right) \times \frac{\text{Annual Percentage Rate}}{\text{Cycles per year}} \\
&= \text{Average Daily Balance} \times \frac{\text{Annual Percentage Rate}}{\text{Cycles per year}}
\end{aligned} \tag{3.4}$$

The introduction of the average daily balance method stopped the once common practice of daily compounding interest application.

Even though the ADB method is used in the vast majority of cases in various forms (e.g. averaged over two cycles), other methods of assessing interest are for instance the previous balance method or the adjusted balance method. In the previous balance method interest is charged on what was owed at the end of the previous billing period, with no credit given for the current month's payments. This method is very simple but very expensive. In the adjusted balance method, interest is charged on the previous month's balance after subtracting payments.

3.2.4 Grace Period

A common feature of credit cards is that if the balance is paid in full every month no interest will be assessed. This is referred to as a **grace period** or **payment exception**. Typically, only the merchandise balance is granted this courtesy, cash balances and transferred balances are not commonly protected. All accounts start out with the grace period active, will maintain it until the first time not the entire balance is paid down. If after a period of revolving payments, the customer starts paying the balance in full again, the grace period will be reestablished after the first two complete balance paydowns.

tons of examples

3.2.5 Minimum Payment Considerations

legal requirement to amortize over less than ten years The minimum payment due is the amount of payment that is required by the minimum payment due date in each cycle in order to keep or make the account current. Failure to pay the minimum payment at the due date will result in delinquency and possibly other consequences such as authorization freezes on the account. The minimum payment due is subject to regulatory and practical requirements. The Credit CARD Act of 2009 requires it to be amortizing over a reasonable period of time assuming no additional charges are posted. In practice, the composition of the minimum payment due (MPD) most frequently used is:

$$\begin{aligned} \text{MPD} = \alpha_m \quad = \quad & \text{Fixed Percentage of Principal} \\ & + \text{All Interest in Period} + \text{Any Amount Past Due} \\ & + \text{All Fees Applicable} + (\text{Any Amount Over Limit}) \end{aligned} \tag{3.5}$$

The fixed percentage of principal is typically in the order of 1.0-1.5% of outstanding balance. The Interest charges apply only if no payment exception is present (grace period). Fees applicable are fees incurred by customer behavior such as late fees, cash advance fees, foreign transaction fees, etc. but not fees that are annual membership or similar fees. If a customer is late or over-limit, these amounts will have to be included, with over-limit not always being part of the minimum payment consideration such as to avoid customers becoming delinquent even though they made minimum payments at all times. Any over-limit amount will have to be paid down though in order to retain charging priviledges on the card.

> **Side Note:**
>
> It is important to know that the amortization requirement of the regulator assumes the consumer to pay the current minimum payment amount for the rest of the amortization period. That will for all subsequent payments require to contribute *more* than the minimum payment due. If only the minimum payment is paid each cycle, the math changes such that it will take longer (and cost more) to pay down the balance. To determine the time it takes to pay down a balance paying only the minimum payment, we can limit the consideration to the principal amount. The Principal P_i at the end of each cycle i at a required fixed percentage of q is:
>
> $$P_i = P_{i-1} \times (1 - q) \quad , \tag{3.6}$$
>
> which leads to
>
> $$P_i = P_0 \times (1 - q)^i \quad . \tag{3.7}$$

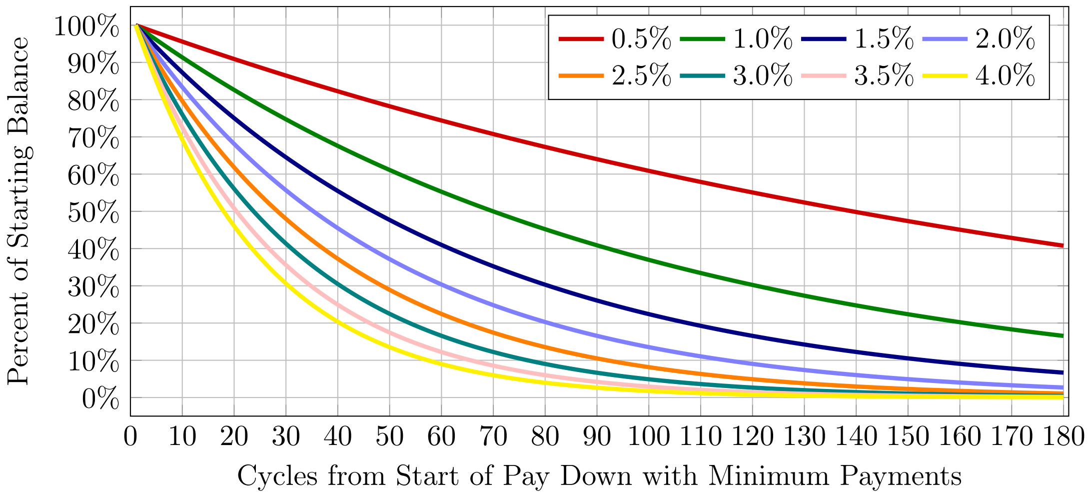


Figure 3.5: Remaining principal as a fraction of original principal balance as a function of cycles for values of q from 0.5-4.0%.

Here we can determine the number of cycles n it takes to get below any threshold amount P_n:

$$\frac{P_n}{P_0} = (1-q)^n \tag{3.8}$$

which results in

$$n = \frac{\log \frac{P_n}{P_0}}{\log(1-q)} \tag{3.9}$$

n needs to be rounded up to nearest natural number.

Example(s):

If P is \$1000, I=2% and q=1%. When would we be owing less than \$1?

$$n = \frac{\log 0.001}{\log 0.99} = -3/-0.004365 \approx 687.3 \tag{3.10}$$

so we would need 57 years to get there. Had we instead used the payment of cycle 1 each month we can use formula 2.36 and obtain

$$\frac{\delta_1}{\alpha} = \frac{q}{q+I} = \frac{1}{3} = (1+I)^{-n} = 1.02^{-n} \tag{3.11}$$

from which we can easily determine n as

$$= -\frac{\log 1/3}{\log 1.02} \approx 0.477/0.0086 \approx 55.5 \tag{3.12}$$

or less than six years which is ten times less than the above computation. If instead I=1%, we obtain $n \approx 70$.

Side Note:

The interest to be paid in each cycle is

$$\epsilon_i = P_{i-1} \times I = I \times P_0 \times (1-q)^{i-1} \tag{3.13}$$

where I is the APR divided by the number of cycles per year.

The overall interest γ_n paid up to a cycle n can then be determined by

$$\begin{aligned}
\gamma_1 &= \epsilon_1 = P \times I \\
\gamma_2 &= \epsilon_1 + \epsilon_2 P \times (1-q) \times I \\
\gamma_3 &= \epsilon_1 + \epsilon_2 + \epsilon_3 = P \times I + P\,(1-q) \times I + P \times (1-q)^2 \times I \\
& \dots \\
\gamma_n &= \sum_{i=1}^{n} \epsilon_i = P \times I \times \sum_{i=1}^{n} (1-q)^{i-1} \\
&= P \times I \times \frac{(1-q)^n - 1}{1-q-1} \\
&= P \times I \times \frac{1-(1-q)^n}{q}
\end{aligned} \tag{3.14}$$

where the geometric series computation as in equation 2.3 was used.

As for the principal δ_i paid in each cycle:

$$\delta_i = P_{i-1} \times q \tag{3.15}$$

which leads to an overall paid principal Δ_n at cycle n of

$$\begin{aligned}
\Delta_1 &= \delta_1 = P_0 \times q \\
\Delta_2 &= \delta_1 + \delta_2 = P_0 \times q + P_0 \times (1-q) \times q \\
\Delta_3 &= \delta_1 + \delta_2 + \delta_3 = P_0 \times q + P_0 \times (1-q) \times q + P_0 \times (1-q)^2 \times q \\
& \dots \\
\Delta_n &= \sum_{i=1}^{n} \delta_n = P_0 \times q \times \sum_{i=1}^{n} (1-q)^{i-1} \\
&= P \times q \times \frac{1-(1-q)^n}{q}
\end{aligned} \tag{3.16}$$

We therefore obtain the overall amount paid up to that point β_n as

$$\beta_n = \Delta_n + \gamma_n = P \times (q+I) \times \frac{1-(1-q)^n}{q} \tag{3.17}$$

Interest of principal, graph vs. several q's , vs. loan payments, etc., overall amount of interest paid

For practical reasons two additional conditions usually apply to the minimum payment due

- a minimum amount α_{min}
- if the balance P is less than the minimum amount α_{min}, the balance is the minimum payment

The minimum payment due is then computed as

$$MPD = \begin{cases} \max\left(\alpha_{\text{min}}, \alpha_m\right) & \text{for } \alpha_{\text{min}} \geq P \\ P & \text{for } \alpha_{\text{min}} < P \end{cases} \tag{3.18}$$

Side Note:

It is easy to determine when the minimum payment is governed by which formula by assessing the principal balance amount threshold P_t that trips the use of α_{min}:

$$\alpha_{\text{min}} > \alpha_m = P \times q + P \times I \qquad \rightarrow \qquad P_t = \frac{\alpha_{\text{min}}}{q + I} \tag{3.19}$$

So if for example $q = 1\%$, $I = 2\%$ and $\alpha_{\text{min}} = \$24$ the resulting P_t is \$800. At \$800 remaining balance the minimum payment will be α_{min}, above \$800 it will be α_m.

disclosure requirements

3.2.6 Fees

Annual membership fees, overlimit fees, late fees, cash transaction fees, foreign currency transaction fees, ...

3.3 Business Considerations

From a business perspective credit cards are a high margin on small credits product.

3.3.1 Managing Credit Card Accounts

Steps:

- Acquire, Activate
- Build balances to encourage revolving
- After 3-6 months on book, use behavioral scoring to
  - Aim line increases at the *meaty middle* - lower middle of behavior score range (higher risk, higher revenue)
  - Control overlimit, cash & delinquent transactions (fee contributions)
  - Segment collections activities by forward roll propensity and treat accordingly

Punitive actions (US) can be

- Causes: universal default - a late on any account , not just the card in question; any delinquency, overlimit
- Reaction: penalty interest rate, raise minimum payment, reduce credit line, close account

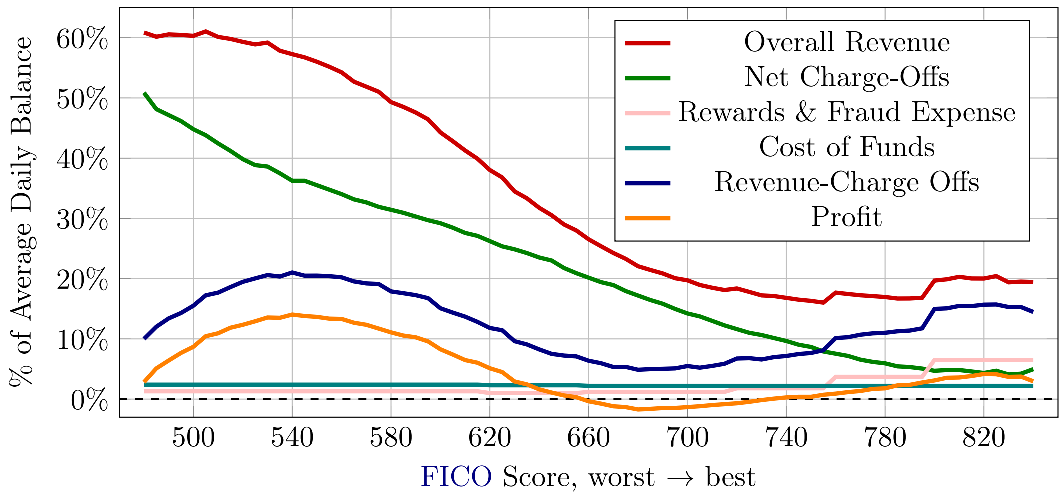


Figure 3.6: Overall revenue, net charge offs, gross margin and profit as a fraction of the average daily balance assuming 3.4% operating and acquisition cost. Data digitized from NBER

3.3.2 Revenue from Cards

From an issuers perspective, credit cards are a high margin product on small credits. Returns on equity of $20^{+}\%$ are typical. The life cycle is relatively short compared to installment loans. It is a good entry point for a wallet and brand building.

The profit for credit cards is of course,

$$\text{Revenue} - \text{Expenses} = \text{Profit} \tag{3.20}$$

Revenue sources can include any and / or all of the following

- Interest Income
- Fee Income
- Interchange Income
- (Insurance Income)

and are influenced by

- Activation Rate / Pricing
- Purchases / Cash Utilization
- Attrition / Retention

The Expenses are manifold and complex as well

- Operating Costs
- Acquisition Costs
- Origination Costs
- Collections Costs
- Servicing Costs
- Funding Costs
- Net Credit Loss
- Charge Offs
- Recoveries

The key loss rate components:

$$\$\ \text{Loss Rate} = \frac{1}{\text{Activation Rate}} \times \text{Incident Loss Rate} \times \text{Balance Control Ratio} \quad , \tag{3.21}$$

where the activation rate is a function of the product / pricing offer, the incident loss rate a function of the score cutoffs and the balance control ratio a function of the credit line.

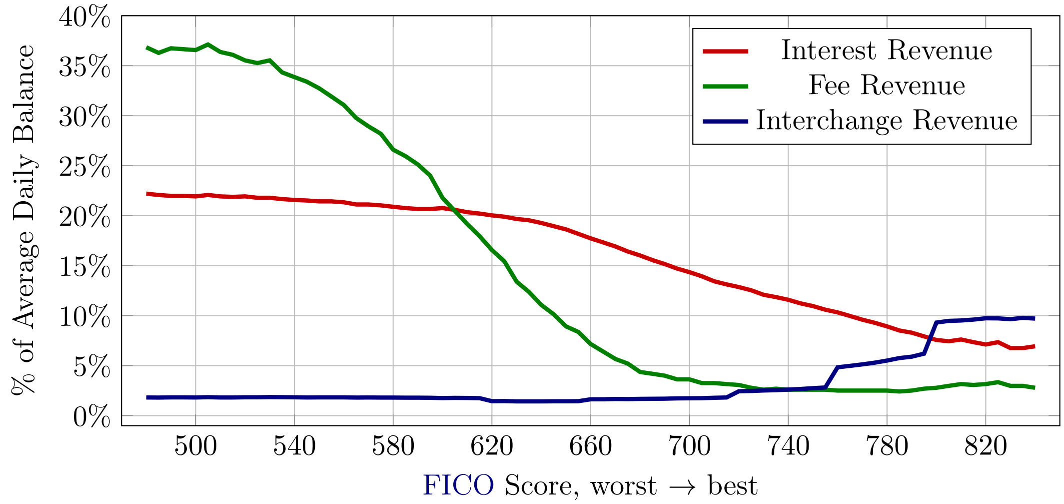


Figure 3.7: Annual revenue by type depending on credit quality / card product. The lower range is dominated by fee revenue which turns irrelevant at about 700 FICO score. Revenue streams are additive meaning, all exist on all cards and need to be summed up, see figure 3.6. Data digitized from NBER

$$\begin{aligned}
\text{\$ Loss Rate} &= \frac{1}{\text{Activation Rate}} \times \text{Incident Loss Rate} \times \text{Balance Control Ratio} \\
&= \frac{\#\text{ Booked}}{\#\text{ Active}} \times \frac{\#\text{ Charge Off}}{\#\text{ Booked}} \times \frac{\text{Avg. C/O Balance}}{\text{Avg. Good Balance}} \\
&= \frac{\#\text{ Booked}}{\#\text{ Active}} \times \frac{\#\text{ C/O}}{\#\text{ Booked}} \times \frac{\text{\$ C/O}}{\#\text{ C/O}} \times \frac{\#\text{ Active}}{\text{\$ Good}} \\
&= \frac{\text{\$ C/O}}{\text{\$ Good}} \qquad (3.22)
\end{aligned}$$

So overall the $ loss rate equals the ratio of charge off balance over good balance. As illustrated in figure 3.6 the net loss rate is very much dependent on credit score. If the limits for all cards in a portfolio are set equal, the balance control ratio will be larger than 1, therefore the money based loss rate is much higher than the loss rate by count as a distinction to well balanced loss rates for installment loans.

Figure 3.6 shows the overall revenue and large group cost components by credit score with the orange line indicating profit. There is a region of credit scores, where the banks in the sample have not been able to turn a profit.

In order to derive the revenue assumptions there are 3 approaches, Fee based, Interest based and Interchange based. All cards have a component of each of these, however, depending on credit quality of consumer there are dominant regions where one of the approaches dominates with up to 70% of the revenue.

Fee: Sub-Prime application fee, membership fee, late, overlimit, cash etc.

Lend: Near Prime Finance Charges based on an APR applied to the average daily balance

Spend: Prime Interchange from merchants based on dollars transacted

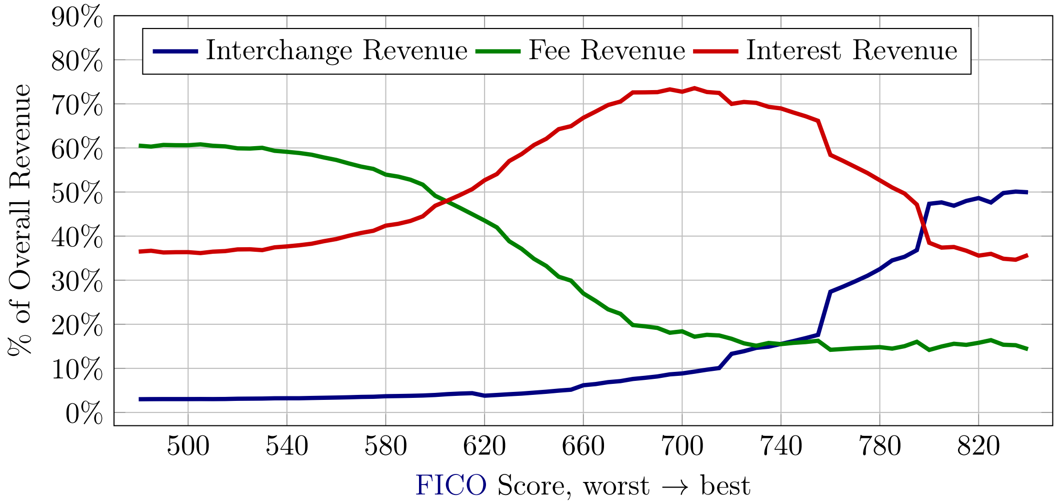


Figure 3.8: Revenue contribution by credit score of all revenue. It is clearly possible to discern the three areas of dominance of one revenue type by credit score range: Fee, Lend, Spend. Data digitized from NBER see text.

Figure 3.7 shows the incomes as a fraction of average daily balance. It can be clearly seen that fee revenue for instance dominates the lower FICO score range and becomes less relevant for higher score consumers. Figure 3.8 relates the contributions of each group to the overall revenue rather than the average daily balance clearly illustrating the above distinctive groups of consumers.

For each model, Fee, Lend & Spend a designated strategy recommends itself

- **Lending** business model
  - very intuitive - charge interest on outstanding balance
  - 3 categories of account balances: Purchase, Balance Transfer and Cash
  - Frequently teaser (introductory) rates on purchases & balance transfers to build balances
  - Interest rates can be determined several different ways: Prime rate + X, Fixed or variable
  - Overall the business success depends on customers **building and maintaining high balances**
- **Spending** business model
  - targeted at customers with high disposable income
  - Based on the Purchase interchange revenue
  - Issuers offer part of the interchange revenue in reward form such as cash back, miles, reward points in order to encourage spending more. As can be seen in figure 3.6 the rewards tend to become a larger share of cost with better credit scores to the point where the larger part of interchange revenue is paid out to customers.
  - The business success depends on getting the customer to **spend more**

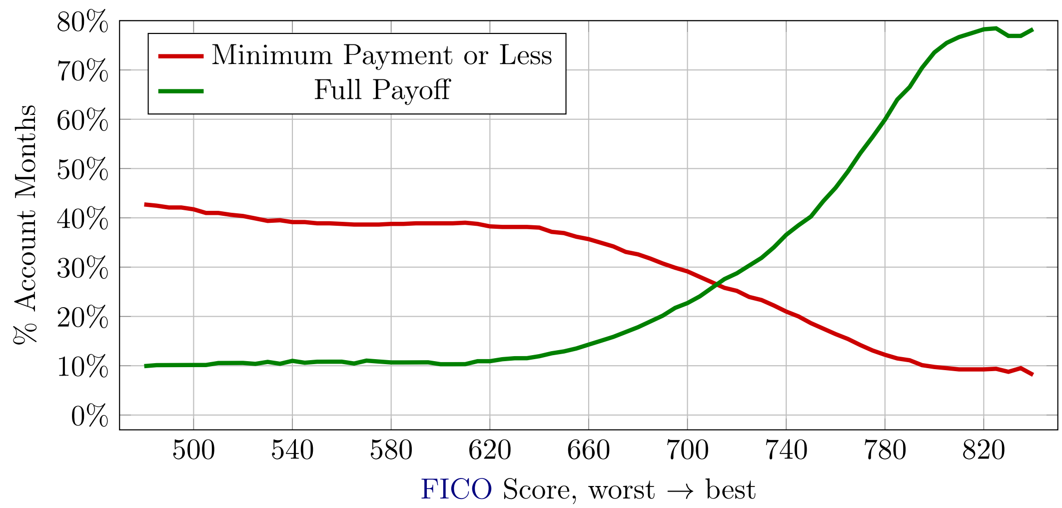


Figure 3.9: Share of account months that pay the minimum payment or less (full revolvers) and those that pay off fully (transactors). There is a strong dependence on credit scores. Data digitized from NBER see text.

- **Fee-based** business model
  - Targeted at customers with few credit options (lower credit scores)
  - Lower card limits, less than $2k, more frequently less than $500
  - Issuer charges fees for specific card benefits: Application, Annual Membership, Cash Advance, Overlimit (regulated by Credit CARD Act of 2009; since irrelevant), Late, Returned check fee
  - Customers pay for issuer flexibility when they exceed limits, pay late, or have other issues
  - Business success assumes customers are **inconsistent with payments**

Figure 3.8 gives a rough idea in which credit score range which strategy would be best applied. Additional data is given in table 3.6 clearly illustrating the different behavior of various customer groups (FICO score buckets). It is important, however, to keep in mind that the data set contains account, credit and performance information of eight large banks between 2008 & 2010 representing roughly 30% of active credit card accounts in the US in that time period which is prior to the Credit CARD Act of 2009 taking effect. While overall trends should hold, specialized offers / lenders can extent ranges / make specific groups more profitable and so forth.

Other segmentation for consumers in a credit card portfolio can be for example:

New/Young: Short time on books, usually younger consumers

Transactor: Customers who transact and pay in full every month. Target these consumers with the Spend model. Figure 3.9 gives an idea on the proportion of transactors among the population by credit score. Super prime credit scores are clearly dominated by transactors.

Revolver: Customers who carry a balance from month to month and generate interest to issuer. Target with Lend model.

High Value: high revolving propensity

Convenience Revolvers: 1-2 times per year - holidays, big purchases

Convenience Users: inactive, but for 1-2 times a year

Inactive: Customers who do not use their cards

Rate Surfers: Customers continually seeking promotional rates

VIP / Private Bank: Consumers with large deposits / other relationships

Overlimit: Customers who have gone over contractual limits

Delinquent: Customers who are late on their payments

Table 3.6: The summary statistics by risk group for NBER review paper. The relative shares and percentage values marked with an asterisk have been added to illustrate the behavior past the original data.

		FICO Score Range					
	Total	<620	620-659	660-719	720-759	760-799	800$^+$
Share of Accounts (%)	100.0	17.3	12.6	24.6	18.6	19.2	7.6
Panel A: Capacity and Utilization (annualized $ per account)							
Credit Limit	8,042	2,025	3,546	7,781	11,156	12,400	11,390
Average Daily Balance	1,410	804	1,469	2,029	1,797	1,110	486
Purchase Volume	1,820	730	1,019	1,651	2,306	2,892	2,282
* Credit limit share (%)	100.0	4.4	5.6	23.8	25.8	29.6	10.8
* ADB share (%)	100.0	9.9	13.1	35.4	23.7	15.1	2.6
* Purchase Volume Share (%)	100.0	6.9	7.1	22.3	23.6	30.5	9.5
* Utilization (%)	17.5	39.7	41.4	26.1	16.1	9.0	4.3
* Purchases of ADB (%)	129.1	90.8	69.4	81.4	128.3	260.5	469.5
* Purchases of Limit (%)	22.6	36.0	28.7	21.2	20.7	23.3	20.0
Panel B: Realized Profits (% of Average Daily Balance)							
Total Income	25.0	45.7	31.5	21.0	16.9	17.1	19.9
Interest Charges	14.3	20.6	19.2	15.2	11.8	9.3	7.6
Total Fees	7.6	23.3	10.9	4.1	2.5	2.4	2.9
Interchange Income	3.2	1.8	1.5	1.7	2.6	5.4	9.5
Total Costs	23.4	37.8	30.2	22.5	17.2	15.6	16.8
Net Charge-Offs	15.6	30.8	23.4	15.8	9.7	6.3	4.7
Cost of Funds	2.3	2.4	2.3	2.2	2.2	2.2	2.2
Rewards and Frauds	2.2	1.3	1.0	1.2	1.8	3.7	6.5
Operational Costs	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Collection	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Marketing + Acquisition	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Other operational cost	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Realized Profit	1.6	7.9	1.3	-1.6	-0.2	1.5	3.1
* Profit of Limit (%)	0.3	3.1	0.5	-0.4	-0.0	0.1	0.1
* Profit of Purchases (%)	1.2	8.7	1.9	-2.0	-0.2	0.6	0.7
* Share of Income (%)	100.0	18.0	16.5	29.7	16.0	10.3	2.1
* Interest (%)	100.0	14.2	17.6	37.6	19.6	9.8	1.4
* Fees (%)	100.0	30.2	18.8	19.1	7.8	4.8	1.0
* Interchange (%)	100.0	5.5	6.2	18.8	19.3	25.5	7.8
* Share of Cost (%)	100.0	15.9	16.9	34.0	17.4	10.1	1.9
* Net Charge-Offs (%)	100.0	19.5	19.7	35.9	14.7	6.1	0.8
* Rewards and Fraud (%)	100.0	10.2	5.7	13.4	15.2	32.3	22.5

3.4 Reference & Data

Reference & Literature:

- Borrower Risk Profiles, Consumer Financial Protection Bureau, May 2017,
- Regulating Consumer Financial Products: Evidence From Credit Cards, S. Agarwal, S. Chomsisengphet, N. Mahoney, J. Stroebel, National Bureau of Economic Research Working Paper # 19484 . The Data shown is from a time period of 2008-2010 prior to the CARD act taking full effect and covers major banks that amount for ≈30% of credit cards at the time, but overall trends are illustrative and should hold for the entire spectrum.
- Wikipedia: Revolving Credit
- Encyclopedia Britannica : Credit Card
- The Consumer Credit Card Market, CFPB, December 2015
- Nilson Report issue 1109, May 2017
- ValuePenguin.com

Chapter 4

Credit Reports and Scores

Credit reports and scores are tools to assess the risk of default / delinquency often referred to as the **willingness to pay**. Lenders use them to determine

- whether or not a loan can be extended to the applicant,
- at which risk of not getting the money back
- and the APR range that would mitigate that risk at which the applicant is able to borrow money.

It is important to know however though, that credit reports and subsequently credit scores **do not** determine and are useless for

- the minimum and maximum loan amounts,
- the minimum and maximum monthly payments the customer can afford or
- the length of the term of the loan.

The latter are subject to business rules as well as legal restrictions (typically on the state level) and may depend - usually do - on factors not related to reporting such as your current income, loan amount to net income ratio etc. such as to determine your **ability to pay**. This is typically dealt with in an *affordability assessment.*

Credit reports and scores have short comings or blind spots that we should be aware of. Intrinsically, they contain no data to account for

- macroeconomic effects such as recessions or boom periods,
- significant changes in customer attitude,
- shifts in local economic developments or
- the effect of adverse selection[1] due to competing products or offerings.

It is important to note, however, that credit reports contain information of the past behavior of an **individual consumer** (the financial **history**), while credit scores are generated by identifying **groups of consumers** with similar characteristics in their

[1] Adverse selection is a gap in the credit policy where the better customers of a class of consumers are offered better conditions by a competitor and what is left is a group of consumers that do not qualify for the competitors business rules. As an example, say Lending Club offers customers at 660 FICO 10% rate loans if their income exceeds $15k, whereas AVANT offers loans at 20% at incomes of $10k+ in that range.

Credit Report

- **individual** customer credit **history**
- **known** past behavior
- trade line data
- public records

Credit Score

- **ranks groups of customers** with similar reports
- prediction / **forecast of future** behavior
- quantitative **odds** lead to **probability**

Figure 4.1: Key differences between credit reports and scores.

credit reports, that is, their past financial behavior and **rank** their **risk of delinquency** in a quantifiable fashion. A summary is given in figure 4.1.

All consumers with similar characteristics should - ideally - have the same credit score range assigned. On the other hand, there may be several groups of characteristics leading to the same risk assessment or credit score range. Inherently, Credit scores are **forward-looking** tools for **predicting** a consumer's credit performance in the **future**.

Ideas to derive a competitive advantage from could be but are not limited to:

- **Compute risk better:** make a better risk assessment than competition based on information in credit reports
- **Use more / combine data:** make a better risk assessment with information not contained in credit reports / scores that is legal to use
- **Offer better conditions:** consequently offer lower APRs due to better, more precise risk assessment
- **Make better decisions** based on data at hand, modify business rules within legal restraints to optimize revenue.

At **AVANT** we try and establish all of those in addition to smoothing out the bumps in the process of obtaining a loan.

A general overview on how the interaction between lenders and credit bureaus, reports and scores work is given in figure 4.2.

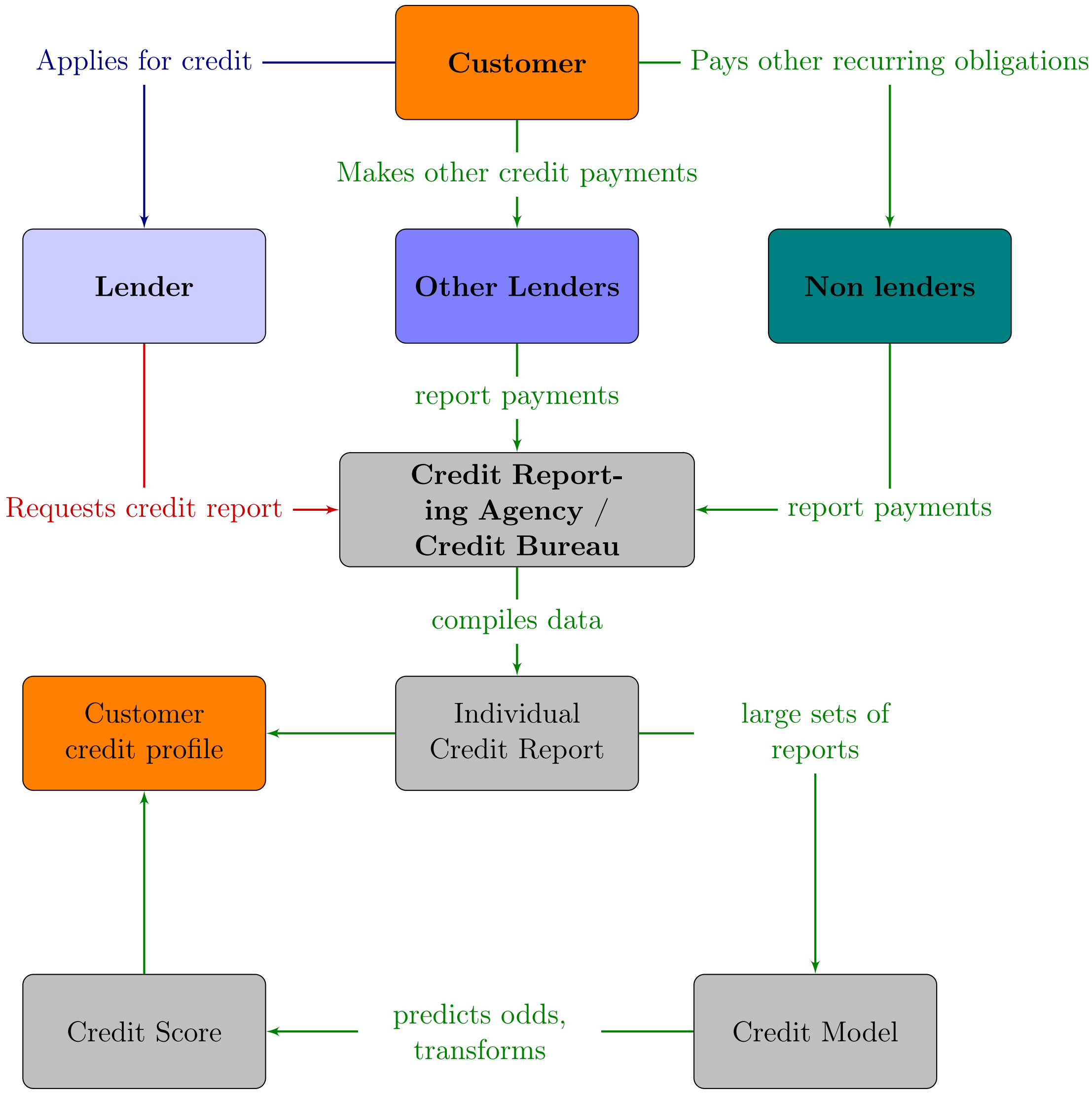


Figure 4.2: Basic credit reporting flow chart. The lender uses the customer credit profile which summarizes his financial history and quantifies the outlook on future payment behavior via a score to assist in the lending / underwriting decision. Developed from Vantage .

The information available for a lending decision can be summarized in three large groups of criteria

- raw data on trade line transactions, public records and inquiries,
- summary variables that group sets of above information and
- credit scores aimed at predicting risk.

Which one or combination of criteria is used in the underwriting decision depends on the sophistication level of the lender. It is common to have a few select criteria in combination with a score, less but increasingly common to have a model based on a large number summary variables or features engineered from raw information. For an individual consumer, the raw data is available as is typically a credit score when purchased, the summary variables are not typically provided to individual consumers.

While the type of information in a report and the relative influence on computing a score remains relatively standardized and therefore unchanged, each credit reporting agency may have business relations with different or the same lenders which results in reports of various agencies possibly containing different information which can result in different credit scores by reporting agency for the same customer. Differences can be local affinity, history, etc.

Table 4.1: The five information zones contained in a credit report with sample values on the example of a popular cartoon character.

IDENTIFYING PERSONAL INFORMATION

Bartholomew JoJo Simpson 742 Evergreen Terrace Springfield, USA 77777	123 Fake Street Springfield, USA 77778 Date of Birth: 4-19-1987 SSN: 123-45-6789	Bart's Employer Bart's Job

PUBLIC RECORD (LEGAL ITEMS)

9-07 Judgment $1,000	Satisfied 3-08

COLLECTION ITEMS

7-06 Collection $500

TRADE LINE (ACCOUNT) INFORMATION

Industry	Date Reported	Date Opened	High Credit	Balance	Current	Historical rating
Bankcard	2-11	3-07	$4,000	$1,200	Current	120+, 6 yrs
Auto Loan	2-11	7-08	$9,500	$1,500	Current	
Retail	12-10	6-04	$1,000	$200	30 days	

INQUIRIES

Date	Industry	Date	Industry
2-12-11	Education	5-06-10	Auto Finance
9-15-10	Energy	10-10-10	Retail

4.1 Credit Reports

Conventional credit reports contain information the full history on trade lines used for borrowing money such as credit cards, personal loans, mortgages and car loans, home equity lines of credit, searches for credit products (inquiries), delinquencies and collections activity or lack thereof. Typically, additional aggregated variables are provided as a summary as well. The historic data can demonstrate responsible handling of credit as well as the opposite.

Credit reports also contain information on public records such as bankruptcies, foreclosures, child support, tax liens and civil judgments. The presence of any public records is an indication of distressed financial behavior and detrimental to the lending decision.

The reports are provided by credit reporting agencies - or bureaus - such as TransUnion , Experian and Equifax, based on information provided by prior lenders, and are used by various institutions, employers and most importantly in respect to **AVANT** , lenders.

Adventures in Education has a very good resource on conventional credit reports. It is very instructive and therefore recommended to explore for a few minutes.

I have stored commented example credit reports for Transunion, Experian and Equifax at the Avant Dropbox.

Table 4.2: Attributes used in the vantages score are designed to represent behaviors on the following dimensions.

Record type	Trade Industry	Behavior Type	Function
Trade	First mortgage	Payment Status	Average
Inquiries	Home equity line of credit	Age	Sum
Collections	Home Equity Loan	Balances	Worst ever
Public records	Installment	Utilization	Highest
Judgment	Auto	Available Credit	Lowest
Bankruptcy	Personal Installment	Account Status	Newest/Oldest
Tax Lien	Student loan		Presence of
	Revolving Credit		
	Bankcard		
	Secured		
	Unsecured		
	Retail		

Side Note:

A very important thing to know is what information a credit report does **not contain.** There can by law not be any information regarding Race, Color, Religion, Nationality, Sex, Marital Status, Age, Salary, Occupation, Title, Employer, Employment History, Where you live, your net worth, assets or your shoe size. In consequence, credit scores are also blind to any of the above. There are legal implications to implicit discrimination therefore all scores and models need to be tested for that.

Usually, credit reporting is done by a lender in a batch process which means that the lender specific updates to a report will occur in regular intervals and not necessarily instantaneous. This can be a source of discrepancy on top of the different reporting bases that the individual bureaus might have.

In the last few years, additional reports have emerged that can be used in the risk assessing computations. The Pay Rent, Build Credit (PRBC) set of data for example extends to all kinds of bill paying without actually borrowing money such as among others rent, electric and other utility bills, phone bills, subscriptions to cable / satellite, internet, rent to own, memberships, retail purchases, prepaid cards, bank accounts and social. This data is a proxy for credit behavior and should be treated as such.

Summary variables

Summary variables provide insight into the report data. They can be used for the design of credit scores. A combination of attributes used for the design of the Vantage score version 3 is given in table 4.2. In addition, trended variables[2] and behavioral characteristics[3] are among the most insightful pieces of information on future outcome.

[2] Any attribute as of the time of application in relation to a time period prior, e.g. your revolving balances today as a percentage of your revolving balances six months ago.

[3] Anything that denotes a specific type of behavior, e.g. do you pay additional principal with each mortage payment, do you keep a revolving balance vs. paying down the enitre credit card statement balance each month?

put in example tradeline reporting

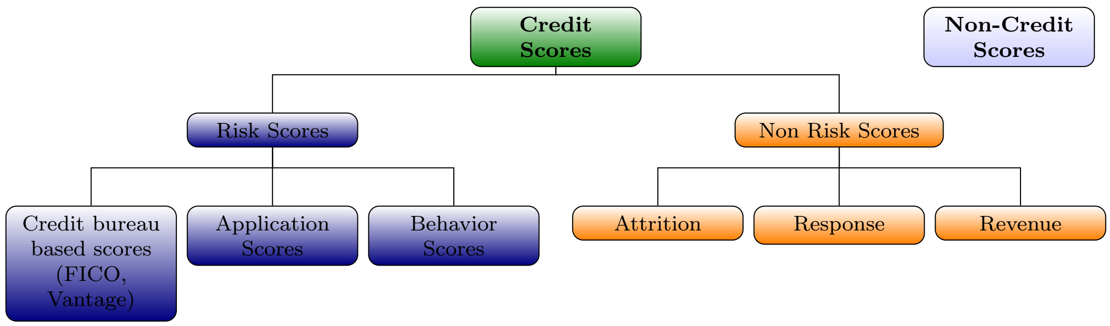


Figure 4.3: Hierarchy of credit scores. From FICO credit boot camp. In common usage, the term *credit score* refers to the bottom left type of scores, the credit bureau based risk scores.

4.2 Credit Scores Basics

Credit risk scoring is a statistical process that converts information from

- **Credit Applicants** and
- **Existing Account Holders**

into a numerical score. This score is regarded as a measure of the credit risk (i.e., the probability of repayment) Source: FICO credit boot camp.

There are a great variety of scores used in daily live of a lender that can be classified into several groups: ***Non-credit Scores*** and ***Credit Scores***. Credit scores can be further narrowed down to **Risk Scores** such as *Credit Bureau Based Scores (credit scores)*, *Application Scores* or *Behavior Scores*, and **Non-risk scores** such as *Attrition*, *Response* or *Revenue scores* as shown in figure 4.3.

There are several key definitions used in the process

Scorecard is a mathematical model used to assess credit risk

Odds are the ratio of one event to another, in credit scoring it is the ratio of goods to bads and measures risk

Characteristic is a question to the credit bureau report such as the *number of bankcards* or *the number of 90+ day delinquencies*

Attribute the answer (value) given by a credit report such as *number of bankcards = **3*** or *the number of 90+ day delinquencies = **2***

Observation date also refered to as Scoring date; the snapshot of bureau data for a defined population used in score development at the beginning of credit activity

Performance date the bureau snapshot for the accounts after the outcome period.

Credit scores are effectively a summary of a consumer's credit file. The odds quote corresponding to a score is representing the general population and not necessarily the applicants of a specific lending product / lender, however, they will still rank order. Most commonly used is the FICO score. It is a three digit number between 300 and 850 which

Table 4.3: Example Partial Scorecard from FICO's Alternative Credit Webinar.

Characteristic	**Attributes**	**Points**
Number of bankcard trade lines	0	15
	1	22
	2	30
	3	40
	4 or more	30
Number of trades with balances >0	0-1	65
	2	55
	3-4	50
	5-7	40
	8 or more	30
Number of months on file	below 12	12
	12 to 23	35
	24 to 47	60
	48 or more	75
Number of months since most recent bankcard opening	0 to 5	20
	6 to 11	25
	12-17	30
	18-23	38
	24 or more	45
Number of months since most recent derogatory public record	No Derog Public record	75
	0 to 5	10
	6 to 11	15
	12 to 23	25
	24 or more	45

rank orders consumers according to risk. Higher scores equate to a lower risk of future default. The Vantage score works in a very similar fashion.

Side Note:
There is no credit score that equates to a zero risk of the bad event in a population. The interpretation has to go along the lines of:

- at score 620 there will be 3 goods for each bad,
- at 640 5 goods for each bad,
- at 660 10 and
- at 680 20.

Delinquencies do happen in each score range but are a lot less likely to occur in the higher score ranges than they are in the lower ones. This is the essence of ranked risk.

A scorecard to determine the credit risk score will be developed from the bureau snapshot of the population at the observation date to generate a prediction of the bads at the performance date. An example partial scorecard is shown in table 4.3. At FICO 08 there are 8 non-derogatory scorecards and 4 derogatory scorecards that use a core

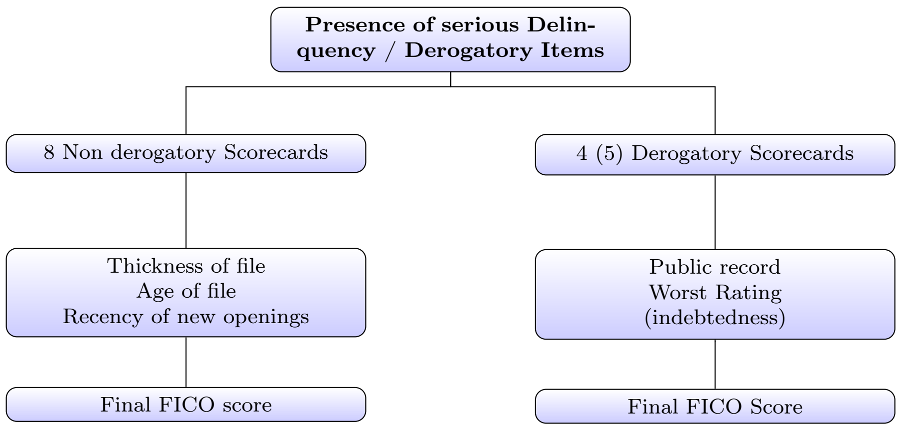


Figure 4.4: Segmentation of Scorecards for FICO 8 (9) Scores starting from the information in a credit bureau file.

group of the 40 most predictive characteristics out of a library of 800+. FICO 9 has an additional derogatory scorecard based on indebtedness. A schematic of the scorecard segmentation for FICO scores 8 and 9 is given in figure 4.4.

add the yes / no nodes

Credit Scores are compiled **solely** from the information filed in credit reports for the purpose of assessing the risk of defaulting on the financial endeavor that is being applied for. They are, in any way, shape or form, nothing but a measure of that estimated default-risk.

It is important to remember that a credit score is forward looking, therefore an output of predictive analytics that can substantially vary between different suppliers. The classification of credit reports into groups of consumers of equivalent risk is the core of a credit score. They usually have a time horizon of less than 2 years and should not be relied on to quantitatively predict anything past that time frame even though ranking will typically carry forward.

Many agencies have proprietary algorithms to calculate scores, the most important being the FICO Score and the Vantage Score. While the rough composition and calculation is known publicly, details are a closely guarded, highly coveted business secret.

For the consumer, the score is key to their financial life.[4] The consumer has the right to know his credit score in as much as what was used for the decision making.

In this manual first we shall look at the properties and usability of credit scores then take a closer look at third party scores.

[4] Adventures in Education

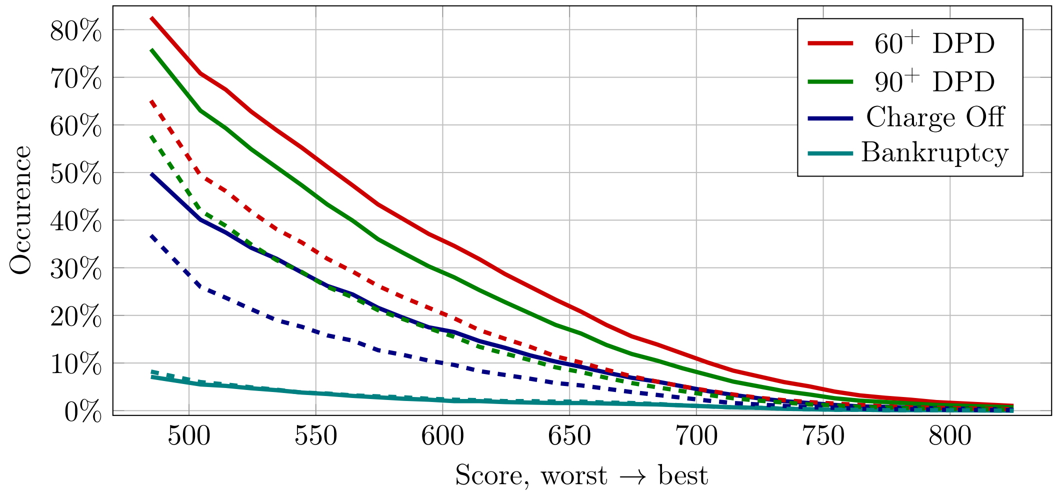


Figure 4.5: Different delinquency targets by FICO 9 score for all account management unspecific to lending product type. The upper curves are super-sets of the curves below (E.g. 60^{+}DPD includes all 90^{+}DPD, charge offs and bankruptcies). The solid lines are for report level data, the dashed ones for trade line level data.

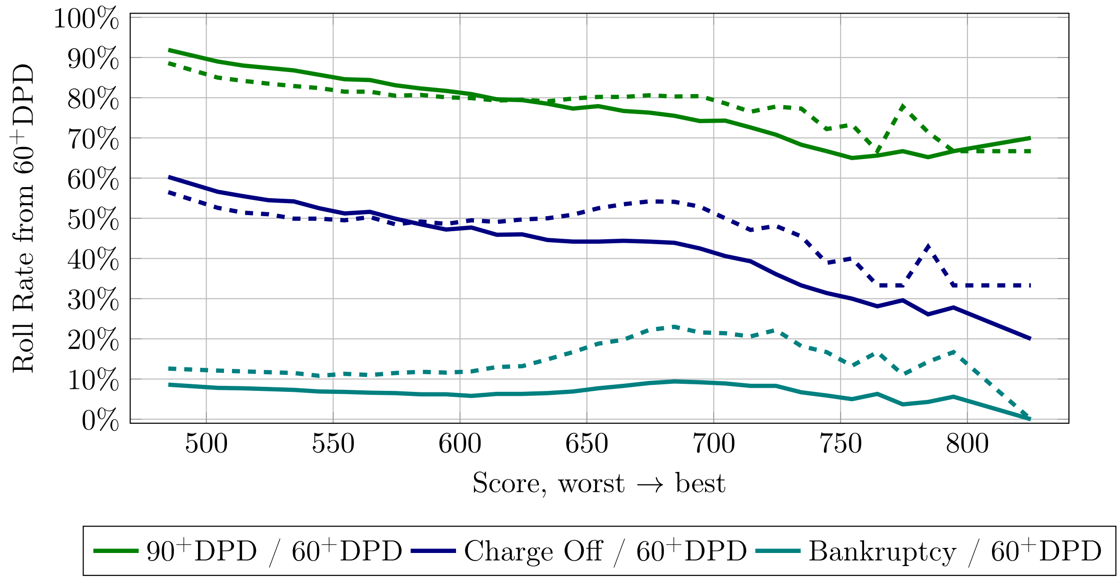


Figure 4.6: Roll rates from 60+ DPD by FICO 9 score for all account management unspecific to lending product type. The solid lines are for report level data, the dashed ones for trade line level data.

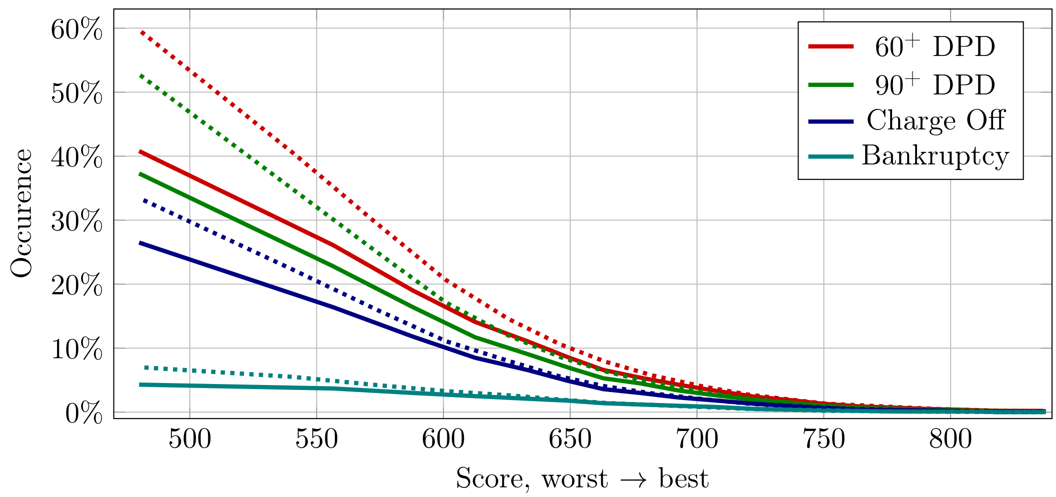


Figure 4.7: Different delinquency targets by Vantage 3 Score for all origination (solid lines) and account management (dotted lines) unspecific to lending product type. The upper curves are super-sets of the curves below (E.g. 60^{+}DPD includes all 90^{+}DPD, charge offs and bankruptcies).

4.2.1 How to characterize a credit score?

In order to use a credit score in an underwriting or subsequently, account management decision, it is of paramount importance to understand its effect and design. A lot of common misconceptions lead to bad or at least less than optimal decisions. With the amount of money at risk in a bad underwriting decision, a thorough understanding of the score is key to performance of the product, especially weaknesses, blind spots, design concepts. For example, if a score was designed to predict a 90 day outcome over a 2 year period, it will most likely not look as efficient for a first installment initial delinquency. So which properties are worth checking / knowing?

- Score design
  - What is the outcome period and target? What is a characteristic event and time frame? Typical are 90 day delinquencies within a one or two year period, filing for bankruptcy or other publicly recorded events. Frequently combinations thereof are used. An example of different characteristic events is given in figure 4.7. Even though the relation does depend on the score range / customer type as demonstrated in figure 4.6, a rule of thumb could be that for each charge off, there will be 1.5x 90^{+}DPD and 2x 60^{+}DPD as well as 0.2x bankruptcies. In other words: Bankruptcies : Charge Offs : 90^{+}DPD : 60^{+}DPD is roughly equal to 1 : 5 : 7.5 : 10.
  - What is the target audience? Is it a general score or is it supposed to predict a specific sector of credit such as car loans, mortgages, credit cards etc. An example of the differences is between different types of credit is given in figures 4.9 and 4.10.
  - Is the score predicting outcome on the account it is used for (account level, newer generations of scores) or a characteristic event on any of the trade lines subject to the score computation (report level)? An example for the difference

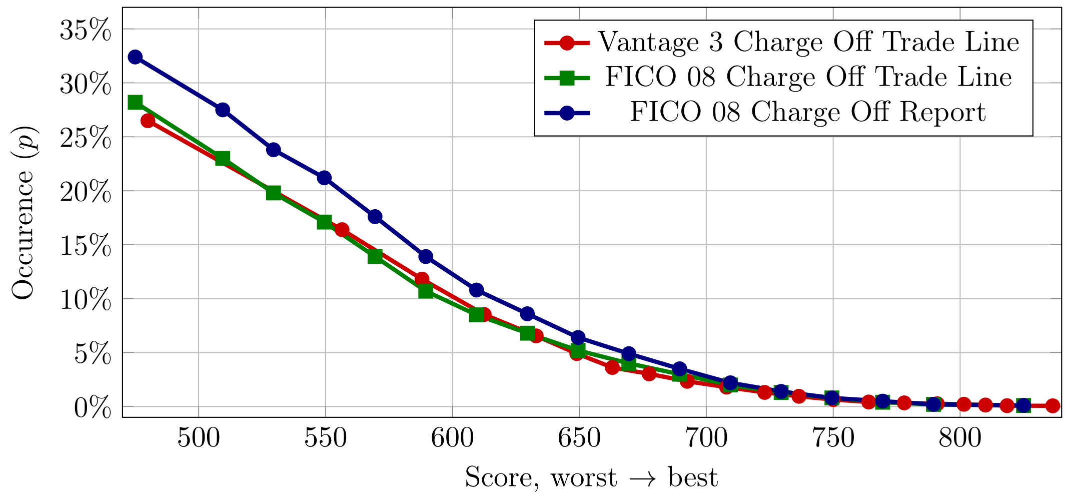


Figure 4.8: Probability of a charge off on a trade line level for Vantage 3 and FICO 08 scores for general account origination. Apparently Vantage 3 and FICO 08 describe the same odds with the same score and are fully comparable ranges at the trade line level. The curve for FICO 08 at the report level is given for comparison.

of a report level outcome vs an account level outcome is given in figure 4.8 and in more detail for the FICO 9 scores in figure 4.5.

- What does the value of the score mean? Typically, scores are based on log odds, meaning the odds of the characteristic event double every x points starting of a base of y. So what is x and y? Typical are a 20 to 1 reference point such as in FICO score or an even odds score where the odds for a characteristic event in the outcome period are 1 to 1 as found in the UK.
- What are the minimum requirements to produce a score? What happens if they are not met? Is the score absent? or a default value?

Underlying data

- Is the performance given for new account origination or account management? An illustration of how different the expected outcomes by score are is given in figures 4.7 for Vantage 3 Scores and 4.18 for FICO.
- Has the score been designed using new or existing accounts? The odds for new accounts vs. existing accounts can easily be different by a factor of 2.
- Which data was used? E.g. was one bureau data or a sample from many credit reporting bureaus? Was there one period of observation or several? If there was one, could it be influenced by external economic developments? As an example, the Vantage Score Version 2 was generated on a sample shortly after the great recession which lead to a heavy over-weighing of recent credit and search activity in the outcome of the score. See section 4.3.3.
- Is the score made up of scorecards, how many are there? How often do individual scorecards get re-calibrated / refreshed?
- Is the score computed from summary variables or underlying trade-line data?
- Does the score itself get re-calibration? If so how and when?

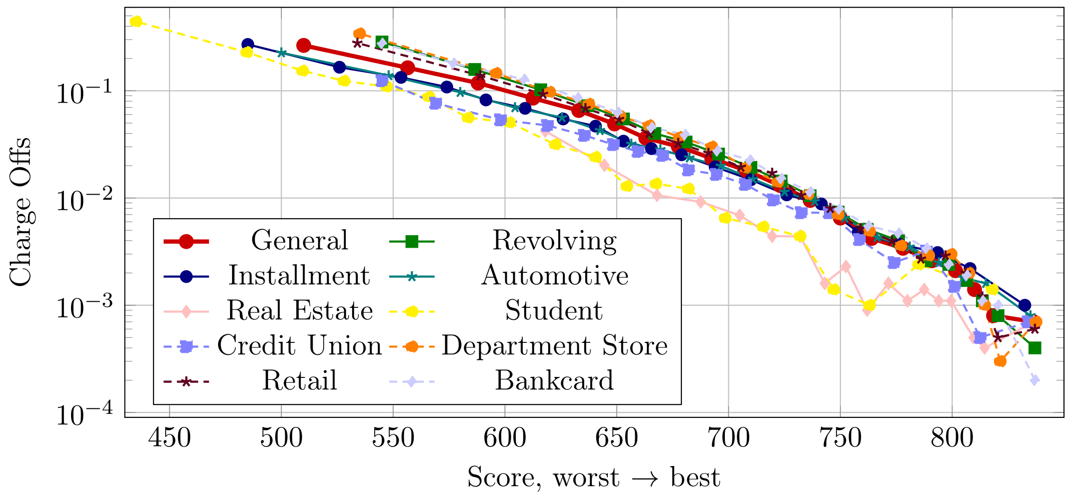


Figure 4.9: Probability of a charge off on a trade line level for Vantage 3 and various account types on a logarithmic scale (the same data on a linear scale is given in figure 4.10). There is a difference in performance of almost one order of magnitude e.g. between revolving and real estate accounts.

- How long and when was the observation period? E.g. Was it before or after the big recession? Does it have any loans originated **online** in its composition?
- How big and what was the population of data used in the score design / validation? Were ther eseveral different observation periods such as to avoid dependence on the position in the macroeconomic credit cycle?
- What defines a good / bad / indeterminate outcome? As an example, very frequently the classification shows

  good the worst status on the account in performance period was no more than 30 DPD, so always paid in full and on time or maximum 1 cycle late

  bad the worst status on the account was 90^{+}DPD or worse (e.g. 90DPD, 120DPD, charge off, bankruptcy etc., also referred to as 3^{+}cycles delinquent)

  indeterminate the worst status was 60 DPD (2 cycle delinquent)
- Is the score produced by machine learning? If so, what kind of models are being used?

Performance Assessment

- A score **must** first and foremost **rank** risk otherwise it is useless. This means do higher scores always perform better than lower scores or vice versa? Does that apply to the full range and if it does not, is the range it does apply the one we want to use? An example is given in figure 4.9 where student loan charge-offs do not properly rank between the scores of 740 and 780 while they do in the remaining ranges.
- How well does the score separate goods from bads (assessed e.g. with the KS metric or, if available, area under curve classification, GINI coefficient)?

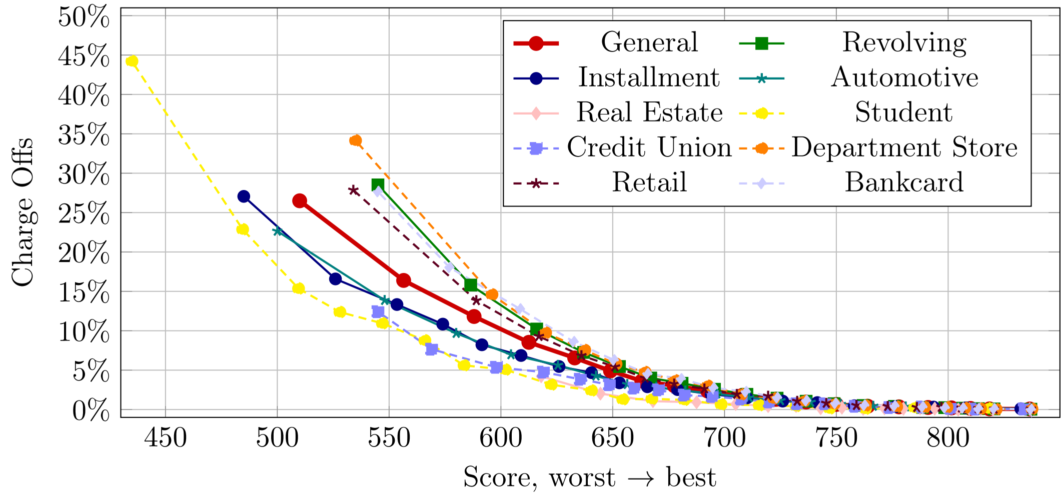


Figure 4.10: This is the same data as figure 4.9 on a linear scale. It shows that real estate accounts require a much higher score to be opened in the first place than most of the other account types. Department store, Bankcard and Revolving accounts perform worse than the general trend. Installment loans and automotive better but close, Credit Unions, Student loans and real estate perform the best vs. score.

- Typically the party responsible for the score creation will provide an odds chart as discussed in section 4.2.3.
- Check distribution vs. population of credit product such as **AVANT** loans
- Check binary target variable: odds of characteristic event in outcome period
- Vary target and outcome period as the underlying ranking may apply to several other constellations
- Check for cumulative losses (money based measures) vs. count (event based measures).
- Always calibrate for the specific situation at hand. An example for the performance of a credit score for different credit products is given in figure 4.9 and 4.10.
- Does it have additional benefits on other variables such as response rate, conversion rate, profiling of customers?
- Does it provide additional ranking versus scores already in use?
- Can it be used to eliminate additional bads? To detect adverse selection?

4.2.2 Probabilities, Ratios & Odds

Credit scores are derived from a linear fit to logarithmic odds of a binary event happening (let's call it a "bad") and it not happening ("good").

$$\ln\left(\frac{\text{goods}}{\text{bads}}\right) = \text{intercept} + \text{slope} \times \text{credit score} \quad . \tag{4.1}$$

If a is the credit score, g the ratio good/bad, i the intercept and $q \neq 0$ the slope we obtain

$$\ln g = i + q \times a \quad \rightarrow \quad g = \exp^{i+q\times a} \quad , \tag{4.2}$$

which transforms into

$$a = \frac{\ln g - i}{q} \quad . \tag{4.3}$$

For easier use, credit scores are typically fixed to a specific ratio of g_0 at score a_0 with defined slope q. We can solve for i

$$i = -a_0 \times q - \ln g_0 \quad . \tag{4.4}$$

Side Note:

Given odds, deriving / producing a score is fairly straightforward, the hard part of the undertaking is, of course, the determination of the odds based on information contained in the credit report. This shall be part of any kind of model, whether it is scorecard based, a regression or tree model. However, knowing their mechanics does prove quite instructive for handling credit scores and designing product(s) using them.

Since humans understand a slope of the form "the odds double every s points" better than natural logarithms or exponentials, it is easier to work the formulas in dual logarithm $\text{ld}\, x = \log_2 x$ rather than natural $\ln x$. So, making b the reference score for a set ratio g_0, a the score for odds g works out to

$$a = b + s \times \text{ld}\left(\frac{g}{g_0}\right) \tag{4.5}$$

and

$$g = g_0 \times 2^{(a-b)/s} \quad . \tag{4.6}$$

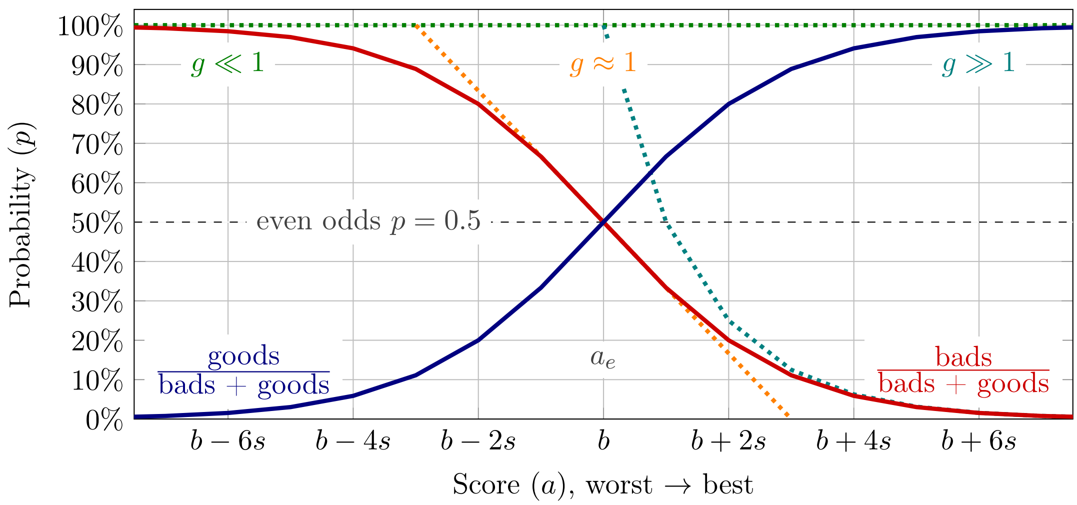


Figure 4.11: Bad and good odds for an example credit score a as a function of b and s in the range of $b - 8s < a < b + 8s$ with $b = a_e$, $c = 1/2$ (reference score is for even odds). Dotted lines are the functional approximations for the limits of g in the respective color.

Obviously, we can derive the formulas for the parameters from the fit.

$$\begin{aligned}
b &= \frac{\ln g_0 - i}{q} \\
b - a &= \frac{\ln g_0 - i}{q} - \frac{\ln g_0 - i}{q} = \frac{1}{q} \ln\left(\frac{g_0}{g}\right) \\
b - s &= \frac{1}{q} \ln\left(\frac{g_0}{2}\right) \\
s &= \frac{\ln g_0 - i}{q} - \frac{1}{q} \ln\left(\frac{g_0}{2}\right) \\
&= \frac{1}{q} \times (\ln g_0 - i - \ln g_0 + \ln 2) \quad .
\end{aligned} \tag{4.7}$$

The probability p of a characteristic event of course is the occurrence of said event in a population.

$$p = \frac{\text{count of characteristic event}}{\text{count all}} = \frac{\text{bads}}{\text{bads} + \text{goods}} = \frac{1}{1+g} \quad . \tag{4.8}$$

Computing odds as a function of p results in $g = 1/p - 1$.

Let a be the credit score, b its base or reference score, c the probability of the characteristic event that the reference score is supposed to predict $c = 1/(1 + g_0)$ and s the step size for which scores double. Then p can be computed as

$$p = \frac{1}{1 + (c^{-1} - 1)\, 2^{(a-b)/s}} = \frac{1}{1 + g_0 \times 2^{(a-b)/s}} \quad . \tag{4.9}$$

Given b, c and s we can determine any credit score a as a function of p as

$$a = b + s \times \operatorname{ld} \frac{p^{-1} - 1}{c^{-1} - 1} = b - s \times \operatorname{ld} g_0 + s \times \operatorname{ld} g \quad , \tag{4.10}$$

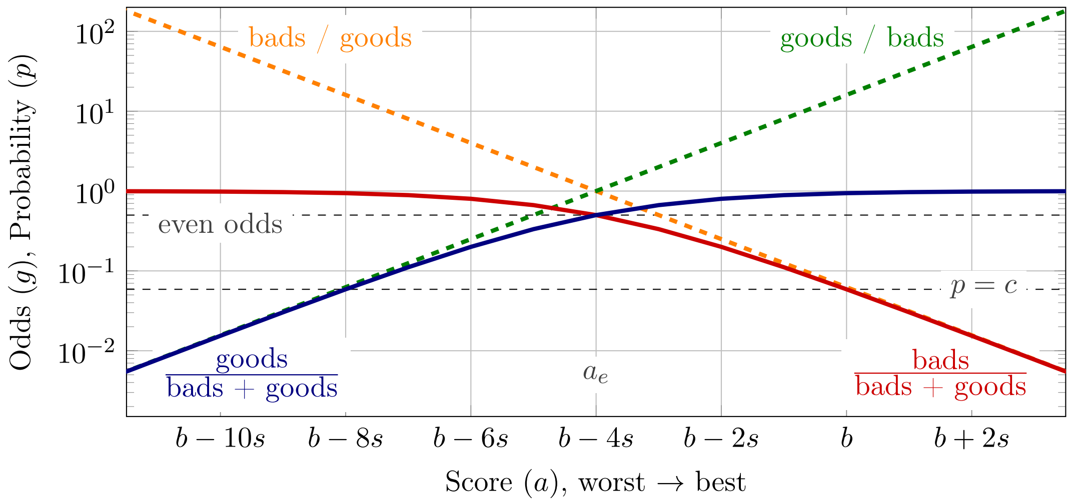


Figure 4.12: Illustration of the bad and good odds on a logarithmic scale for an example credit score a in the range of $b-12s < a < b+4s$ as a function of b and s with $b \neq a_e$, $a_e = b-4s$, $(c = 1/(1+16))$. Additionally, the ratio of bads to goods g^{-1} and vice versa g is shown, demonstrating their doubling / cutting in half every s points.

which for the even odds score a_e for any target probability $0 < c < 1$ reduces to

$$a_e = b + s \times \mathrm{ld}\, \frac{1}{c^{-1}-1} = b - s \times \mathrm{ld}\, g_0 \quad . \tag{4.11}$$

Limits & Approximations:

The behavior for various approximations can be characterized as follows:

$$p|_{g \ll 1} = 1 \qquad \& \qquad p|_{g \gg 1} = \frac{1}{g} \quad , \tag{4.12}$$

for $g \approx 1 \rightarrow a \approx a_e$ the result is

$$p = \frac{1}{1+2^{(a-a_e)/s}} \approx \frac{1}{2+(a-a_e)/s \times \ln 2} \approx \frac{1}{2}\left(1 - \frac{a-a_e}{2s}\ln 2\right) \quad . \tag{4.13}$$

This works particularly well in the range of $a_e - s \leq a \leq a_e + s$ as illustrated in figure 4.11.

For $a = b$, the base score, we find

$$p_{a=b} = \frac{1}{1+g_0 \times 2^{(a-a)/s}} = \frac{1}{1+g_0} = c \quad . \tag{4.14}$$

Side Note:

For $c = 1/2$, or even odds $g_0 = 1$, equation 4.9 simplifies to

$$p = \frac{1}{1+2^{(a-b)/s}} \quad . \tag{4.15}$$

Apparently , we can also back into a score from known p or g, b, $c = 0.5$ and s

$$a = b + s \times \mathrm{ld}\,\left(p^{-1} - 1\right) = b + s \times \mathrm{ld}\,g \quad . \tag{4.16}$$

Given two scores a_1, a_2 with the same outcome period and characteristic it is possible to convert one to the other using their respective parameters $b_{1/2}$, $c_{1/2}$, $g_0^{1/2}$ and $s_{1/2}$

$$a_2 = b_2 + s_2 \frac{\mathrm{ld}\,\left(c_1^{-1} - 1\right)}{\mathrm{ld}\,\left(c_2^{-1} - 1\right)} \times \frac{a_1 - b_1}{s_1} = b_2 + s_2 \frac{\mathrm{ld}\,g_0^1}{\mathrm{ld}\,g_0^2} \times \frac{a_1 - b_1}{s_1} \quad , \tag{4.17}$$

which for $c_1 = c_2$ or $(g_0^1 = g_0^2)$, respectively, simplifies to

$$a_2 = b_2 + s_2 \frac{a_1 - b_1}{s_1} \quad . \tag{4.18}$$

Add in here all the wonderful other things and odds chart provides including cum distributions and KS metric

Table 4.4: Odds Chart Table for a hypothetical credit score. For illustration purposes. Explanation in the text.

Score Band	Bads	Goods	Total	Odds	negative to base (%)	cumulative population (%)	cumulative bads (%)	cumulative goods (%)	KS metric (%)
350-359	733	11	744	0.015	98.47	0.15	0.53	0.00	0.53
360-369	856	16	872	0.019	98.16	0.33	1.16	0.01	1.15
370-379	1,042	23	1,065	0.022	97.83	0.54	1.92	0.01	1.90
380-389	1,283	35	1,318	0.027	97.33	0.81	2.85	0.02	2.83
390-399	1,554	52	1,606	0.034	96.74	1.14	3.98	0.04	3.94
400-409	2,024	66	2,090	0.033	96.85	1.56	5.46	0.06	5.40
410-419	2,119	86	2,204	0.040	96.12	2.01	7.00	0.08	6.92
420-429	2,529	140	2,669	0.055	94.76	2.55	8.85	0.12	8.72
430-439	2,890	162	3,051	0.056	94.70	3.17	10.95	0.17	10.78
440-449	3,152	223	3,375	0.071	93.40	3.85	13.25	0.23	13.02
450-459	3,739	310	4,049	0.083	92.35	4.67	15.97	0.32	15.66
460-469	3,435	428	3,863	0.124	88.93	5.46	18.48	0.44	18.04
470-479	4,540	503	5,043	0.111	90.03	6.48	21.78	0.58	21.21
480-489	4,093	759	4,852	0.185	84.36	7.46	24.77	0.79	23.97
490-499	4,377	841	5,218	0.192	83.89	8.52	27.95	1.03	26.93
500-509	4,733	1,182	5,915	0.250	80.02	9.72	31.40	1.36	30.04
510-519	5,543	1,406	6,949	0.254	79.76	11.13	35.44	1.75	33.69
520-529	5,669	1,864	7,533	0.329	75.26	12.66	39.57	2.28	37.29
530-539	5,264	1,991	7,254	0.378	72.56	14.13	43.41	2.84	40.57
540-549	5,078	2,633	7,711	0.519	65.85	15.70	47.11	3.58	43.53
550-559	5,353	3,253	8,606	0.608	62.20	17.44	51.01	4.49	46.51
560-569	5,179	3,439	8,617	0.664	60.10	19.19	54.78	5.46	49.32
570-579	6,042	4,198	10,240	0.695	59.01	21.27	59.18	6.64	52.54
580-589	5,583	5,250	10,833	0.940	51.53	23.46	63.25	8.11	55.14
590-599	5,196	6,538	11,734	1.258	44.28	25.84	67.04	9.95	57.08
600-609	4,683	6,492	11,176	1.386	41.91	28.11	70.45	11.78	58.67
610-619	4,273	8,332	12,605	1.950	33.90	30.67	73.56	14.12	59.44
620-629	4,677	9,106	13,783	1.947	33.93	33.46	76.97	16.68	60.29
630-639	3,737	10,720	14,458	2.868	25.85	36.39	79.69	19.69	60.00
640-649	3,646	9,349	12,995	2.564	28.06	39.03	82.35	22.32	60.03
650-659	3,466	12,706	16,172	3.665	21.43	42.31	84.88	25.89	58.98
660-669	3,072	10,983	14,056	3.575	21.86	45.16	87.11	28.98	58.14
670-679	2,840	14,578	17,418	5.133	16.31	48.69	89.18	33.08	56.11
680-689	2,293	13,873	16,166	6.049	14.19	51.97	90.85	36.97	53.88
690-699	2,150	13,778	15,928	6.408	13.50	55.20	92.42	40.85	51.57
700-709	1,988	16,435	18,423	8.267	10.79	58.94	93.87	45.47	48.40
710-719	1,715	17,124	18,840	9.982	9.11	62.76	95.12	50.28	44.84
720-729	1,440	16,972	18,412	11.786	7.82	66.50	96.17	55.05	41.12
730-739	1,045	17,204	18,249	16.460	5.73	70.20	96.93	59.89	37.04
740-749	847	16,092	16,938	19.008	5.00	73.63	97.55	64.41	33.14
750-759	694	15,936	16,630	22.973	4.17	77.01	98.05	68.89	29.16
760-769	656	14,275	14,931	21.754	4.39	80.04	98.53	72.90	25.63
770-779	532	16,379	16,911	30.788	3.15	83.47	98.92	77.50	21.41
780-789	382	14,911	15,293	39.029	2.50	86.57	99.20	81.70	17.50
790-799	303	11,851	12,154	39.136	2.49	89.03	99.42	85.03	14.39
800-809	234	13,185	13,419	56.366	1.74	91.75	99.59	88.73	10.86
810-819	214	11,162	11,376	52.177	1.88	94.06	99.74	91.87	7.87
820-829	141	10,993	11,134	78.237	1.26	96.32	99.85	94.96	4.89
830-839	118	9,073	9,191	76.791	1.29	98.18	99.93	97.51	2.42
840-849	94	8,859	8,952	94.729	1.04	100.00	100.00	100.00	0.00
total	137,244	355,776	493,021	2.592	27.84	100.00	100.00	100.00	0.00

4.2.3 Validation / Odds charts

To set any kind of cutoff for a new product requires two key pieces of information:

- Volume distribution: how the applicant population will **distribute**
- Performance: how the new accounts are expected to **perform**

Both of those pieces of information can be found in odds charts.

I have stored the Odds Charts for the most common scores in use on Google Drive (internal only).

A hypothetical odds chart is given in table 4.4. We shall use it as an example for subsequent discussion. Figures 4.13 through 4.15 in this section have been derived from the data in this table. This Odds chart has been generated with a random seed on top of two normal distributions centered at 550 (bads) and 730 (goods) with 100 points standard deviation each to approximate the range that is most familiar for credit scores. The input of good to bad ratio without the random seed was 2.79 goods for each bad.

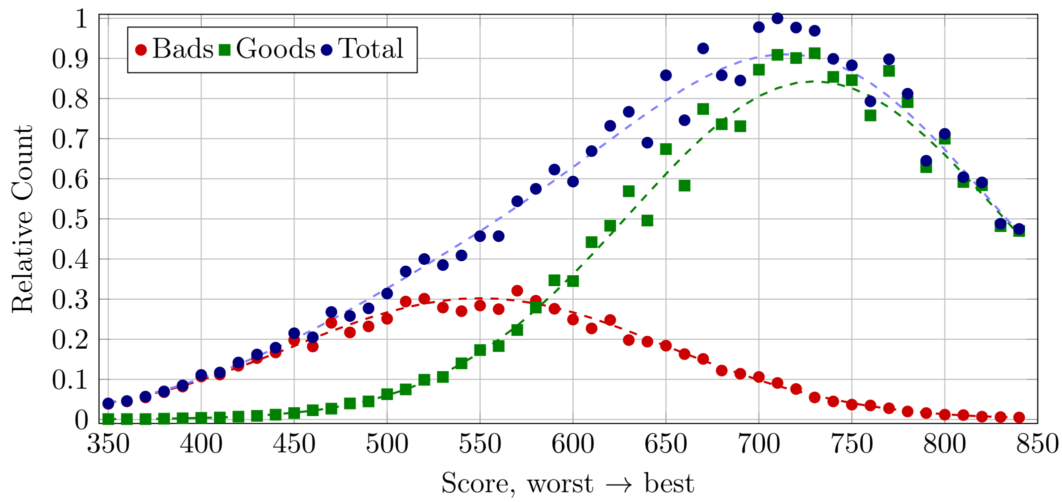


Figure 4.13: Good / Bad / Overall Distributions from table 4.4. Dashed lines for smoothed data.

The average score is 664 and the median at ≈680.

We can easily determine that the even odds score is ≈590. The points to double the odds can easily be determined using equation:

$$s = \frac{849 - 350}{\mathrm{ld}\,(94.73/0.015)} = \frac{499}{12.58} \approx 40 \quad . \tag{4.19}$$

How to understand the data?

An odds chart usually contains a table and a set of parameters used to produce the table. It always has to have a definition of the outcome that is being used to generate a score, which population is included, when was the population observed and when was the performance observed.

As an example, we could look at all accounts generated in June 2012 and observe whether or not they had a severe delinquency of 90DPD or worse until June 2014 when we take stock of those accounts.

To be useful, an odds chart needs to contain the distribution of applicants that got approved by score bucket in general and the distribution of negative performance by score bucket. All necessary properties can be derived from that. Figure 4.13 shows those fundamental bad and good counts by score.

What makes up the individual pieces of an odds chart?

Typically several derived pieces of information are given and / or can be easily derived:

- the remaining information (good or bad count) by score interval *(see e.g. table 4.4)*
- the share of total / bad / good observations in score interval to assess the distribution of scores in a population *(see e.g. figure 4.13)*
- the share of population below (cumulative) or above (descending cumulative) a score bucket *(as shown e.g. in figure 4.15 for the cumulative case)*

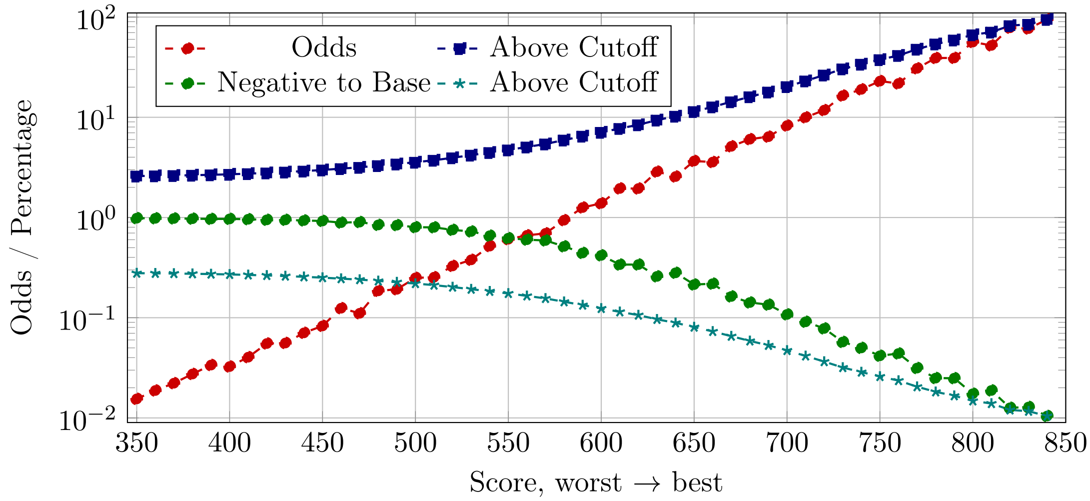


Figure 4.14: The interval good to bad odds, the good to bad odds above the score used for cutoff, interval negative to base outcomes and negative to base outcomes above cutoffs from table 4.4.

- the share of bad / good count below (cumulative) or above (descending cumulative) a score bucket of all bads / goods in the overall sample *(as shown e.g. in figure 4.14)*
- the good / bad rate per interval *(see e.g. the negative to base column of table 4.4 or figure 4.14)*
- the good to bad ratio (the odds) in interval *(E.g. column Odds of table 4.4 or figure 4.14)*
- the odds above or below a specific score band *(E.g. in figure 4.15)*
- the KS metric: cumulative bads percentage minus cumulative goods percentage *(See figure 4.14)*

How to use odds charts to set cutoffs?

There are several options to set the cutoffs based on odds charts

- Cumulative projected bad rate or odds of approved population (above cutoff)
  - *Say we want to have 10 times as many goods above cutoff as bads (for the overall portfolio).*
  - *As can be seen from figure 4.14 we would then set the cutoff score to 650*
- Marginal (interval) bad rate or odds of highest risk approved applicant
  - *Say we want our worst customers to have 10 times as many goods as bads.*
  - *From figure 4.14 as well as from table 4.4 we find the cutoff to be 720 which is substantially higher than in the above case*
- Maximum separation between goods and bads (find score value for maximum KS value=cumulative bads percentage - cumulative goods percentage of population)

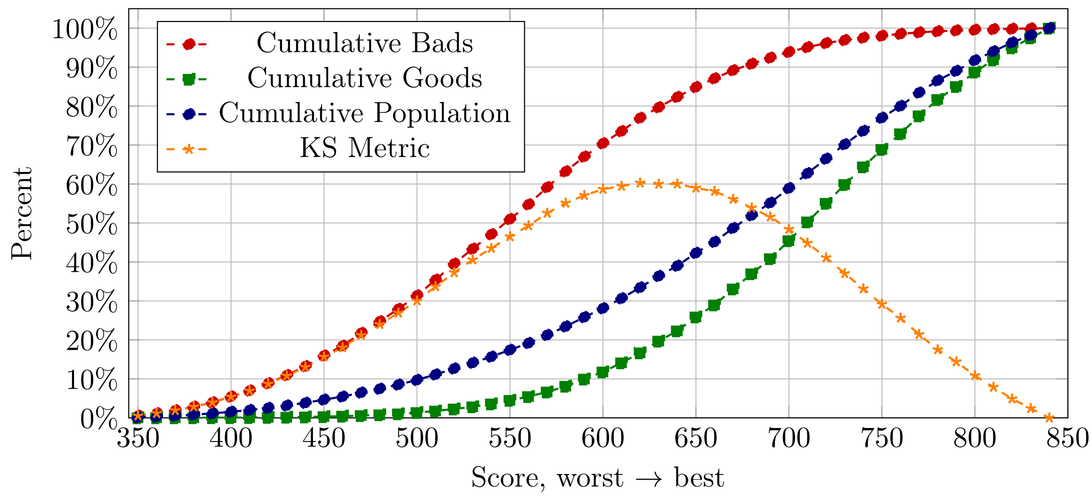


Figure 4.15: KS metric and cumulative distributions from table 4.4

  - *from table 4.4 and figure 4.15 we find the separation is maximal at score interval 620-629 with a value of 60.3%. This is a great value for separation.*

- Breakeven point
  - average good =\$200 revenue, average bad = \$2160 loss
  - breakeven point is where odds of highest risk applicant equal breaking even
  - we need to set the score to the band with interval odds exceeding \$2160/\$200=10.8
  - *in table 4.4 as well as in figure 4.14 it can be easily seen that we need to set the score cutoff to 720 to accomplish this*
- Competition / Regulation dependent
- Based on population ranking
  - *Let's assume we are interested in the top 20% of applicants*
  - *Looking at table 4.4 or figure 4.15 we determine the cutoff required to be at 760.*

4.2.4 What defines score ranges such as *Prime*?

In its most general sense, a *prime* credit score is a better than average score, with a lower than average loss expectation. It is a reflection on the *ranking* property of scores rather than a specific number. Going by this definition and historic charge off rates as shown e.g. in figure 1.10 at the peak of the last recession (2009) a better than 3% charge off rate would have been sufficient to qualify as prime, where as typically sub 1% is required in a normal economic growth period.

A relatively robust definition in 20% quantiles of a ranking of a population by likelihood to go delinquent goes from *deep subprime, subprime, near prime, prime* and *super prime* from the lowest to highest ranked scores or highest to lowest default risk, respectively. The range you are in will depend on the individual lender and specific situation, but overall, lower scores will result in higher rates for credit to be paid and less money issued to the customer up to the point of not extending any credit at all. An illustration of the ranges in an example scenario is given in fig. 4.16.

Prime and the other classifiers for credit quality are a population ranking, *NOT* an actual credit score or probability even though both of these can be deduced[5].

The TransUnion Industry Insights Report gives a distribution for all consumers with a Vantage 3 score in Q4 2016 as[6]

- 29% in the range 300-600 (subprime)
- 40% in the range of 721-850 with 18% in the range 721-780 (prime plus) and 22% in the range 781-850 (super prime)
- 14% in 601-660 (near prime) and
- 17% in 661-720 (prime)

A distribution from FICO is available for the year 2016 and gives base don TransUnion reports a distribution of FICO 8 as:

- 16% in 300-579
- 19% in 580-669
- 23% in 670-739
- 23% in 740-799
- 19% in 800-850

For a consumer, being prime or subprime has significant implications on the general availability of credit and the cost of credit. An illustration is given in figure 4.17. Comparisons between the overall cost of a mortgage by credit quality and associated APRs can be easily found on the internet.

Maybe do a table.

Define Prime, subprime, Rule of thumb delinquency rates for each rating. Go by reason codes. No bankruptcy, high utilization, age of credit line etc. no derogs means prime yada yada yada

[5] A lot of publications will claim that 40% of American consumers have subprime credit scores, which - looking at the above definition - is the very essence of that classification and not at all surprising or reason for concern. What is of concern is that consumers without prime or super prime credit scores are not served by mainstream financial institutions.

[6] Prime level definitions are from TransUnion .

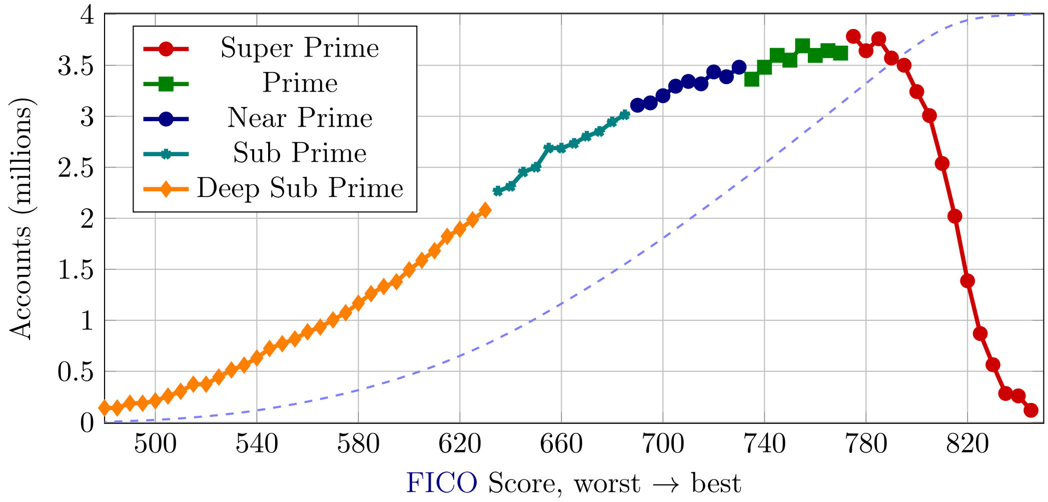


Figure 4.16: Distribution of credit card accounts by FICO score. The division into 20% quantiles is given to determine *super prime, prime, near prime, sub prime* and *deep sub prime*. The median value here can be located in the cumulative plot (dashed) at around 710, the average score in this case is 705, making all customers above ≈720 *prime* in a general sense. Data digitized from NBER

show curve for loans / credit cards over time and score, make transformation count to money, account for timing, refer UE

use odds charts for various fico scores to show the dependence on product of quantiles

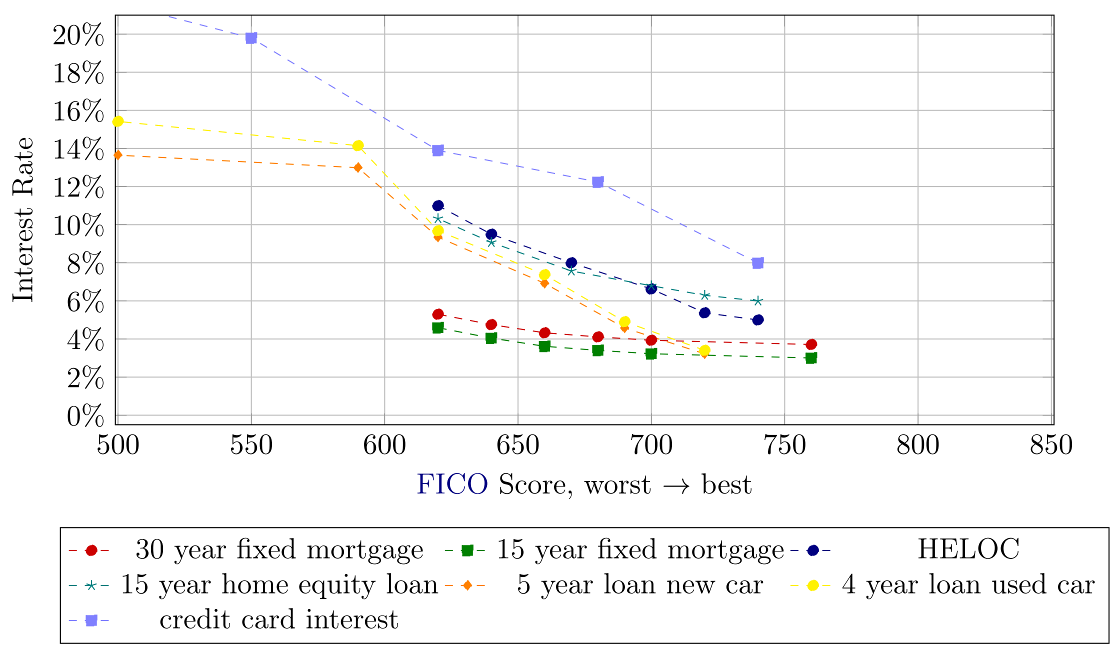


Figure 4.17: Effect of credit quality on interest rate for various credit products. Derived in May 2017 from myfico.com. The interest rate stated is given for the lowest score that qualifies for it.

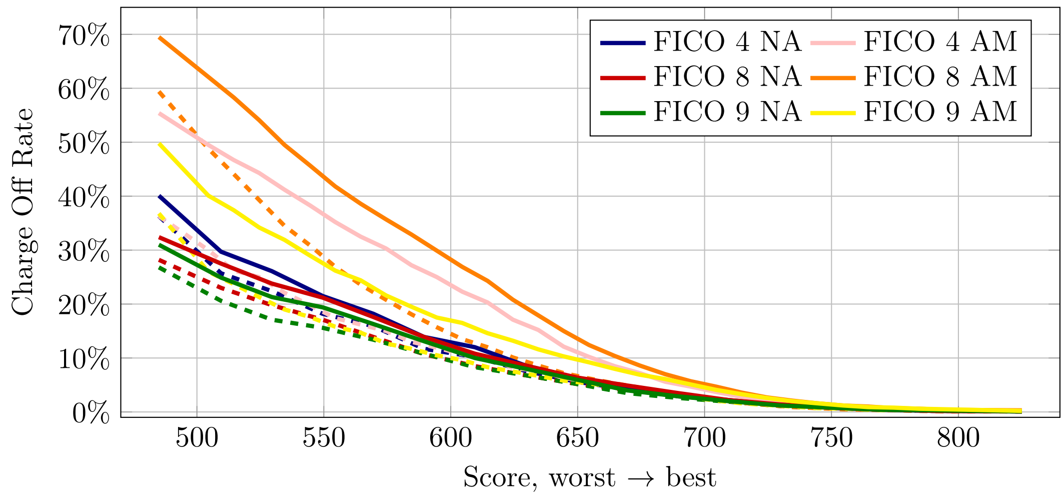


Figure 4.18: Charge Offs by FICO generation for originations (New Accounts, NA) and account management (AM) on a report level (solid lines) and trade line level (dashed lines). The difference between the score generations is much more pronounced for account management and scores below 700 than it is for originations and scores above 700.

4.3 Common Commercial Credit Scores

Third party credit scores have been and currently are dominated by FICO scores and scores produced by the credit reporting bureaus or subsidiaries such as the Vantagescore Solutions, LLC.

The most common credit score is the FICO score that appears in 50+ versions. It is released by the FICO corporation. FICO does produce the scores with a proprietary algorithm based on credit reporting bureau data. FICO itself does not possess or hold any consumer reports. This service is a paid one, the avoidance of which is the less altruistic motive to produce Vantage Scores by the three bureaus. A brief history of the credit scores is shown below.

4.3.1 History

1956 FICO founded in the San Francisco Bay Area

1958 FICO delivers first custom credit scoring system

1970 FICO delivers first custom credit scoring system for a bank credit card

1989 FICO launches first broad-based FICO® Score

1991 FICO® Scores are available at all three major US credit reporting agencies

1995 Fannie Mae and Freddie Mac endorse the use of FICO® Scores for evaluating mortgage loans

2000 FICO introduces FICO NextGen Scores

2001 FICO launches myFICO.com, giving consumers access to their FICO® Score

2004 FICO introduces FICO® 4 Score, still used by GSEs and FHA today

2006 Vantage Score 1 introduced by triple bureau alliance

2009 FICO introduces FICO® 8 Score

2009 FICO introduces FICO® Mortgage Score

2011 Vantage Score 2 introduced

2013 Vantage Score 3 introduced

2014 FICO introduces FICO® 9 Score

2015 FICO alternative data score released **check this**

2017 Vantage Score 4 due for release in October

The scores currently most widely used are FICO 4 in all mortgage lending, FICO 8 is ubiquitous among all other lending products with the exception of a few old school lenders usually smaller ones. The transition to FICO 9 is currently underway, as of June 2017 none of the big lenders has switched to use of FICO 9 even though by design it should outperform the previous scores. Vantagescore 2 has become widely used, currently being transitioned to version 3. Version 3 is what is made available by consumer websites such as CreditKarma as the quoted score. We shall discuss the peculiarities of each score further down in the text.

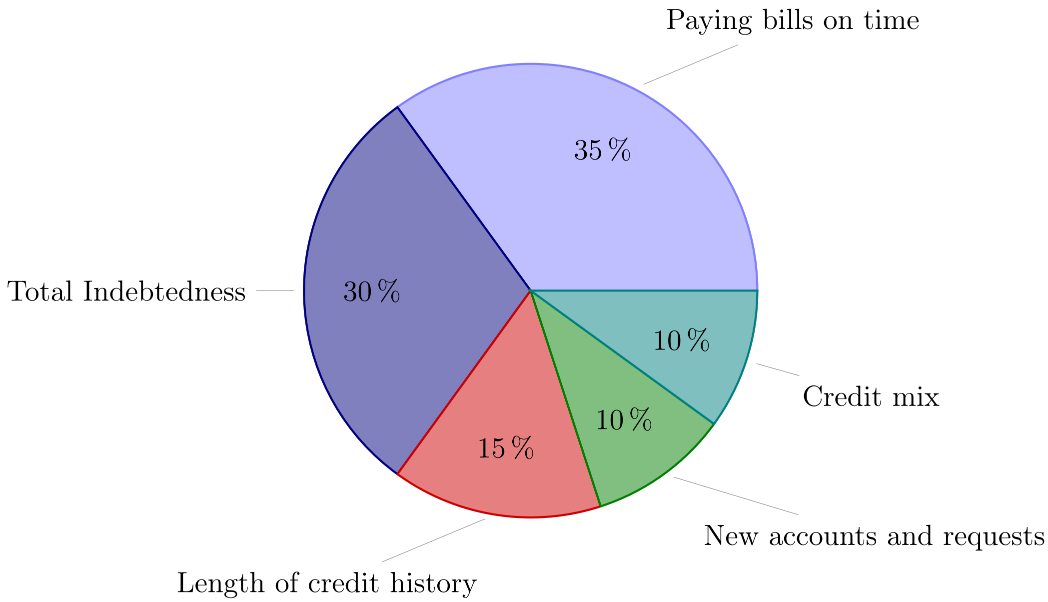


Figure 4.19: Weight of categories in FICO scores according to FICO .

4.3.2 What makes up the individual factors of a FICO score?

FICO scores range from 350 (ridiculously bad) via 500 (very bad) to 850 (extremely good) with a median of 720. For every thirty points the default risk doubles and consumers with FICO scores of below 640 (sub prime) usually find it hard to get credit at favorable conditions from lenders or banks.

While useful as a tool and very conservatively used by banks, the existing credit scoring systems do have shortcomings. For one, negative events such as bankruptcy, late payments or foreclosures stay on record for a very long time. In general, negative events have an impact that far outweighs a history of good financial behavior. See section 4.3.4 for more details. Moreover, there are minimum scoring criteria to produce a FICO Score that are not met by a substantial fraction of US consumers. See section 4.3.5.

Lets look into published details on how each information piece is used

- **Payment history** for 35% at FICO . Did you pay on time previously? A relative dependence of risk on payment history is shown in figure 4.20 based on FICO credit boot camp.
  - Account types investigated include CreditCards, Retail Accounts, Installment Loans, Finance Company Accounts and Mortgage Loans
  - Public records and collection items that will be looked at are all rather negative such as bankruptcies[7], Foreclosures, Lawsuits, Wage Attachments, Liens, Judgments
  - Details on late or missed payments considered are, how late they were, how much was owed, how recently they occurred and how many there are
  - Good records on most accounts has favorable effects on your score

[7] Those stay on your credit report for 7-10 years

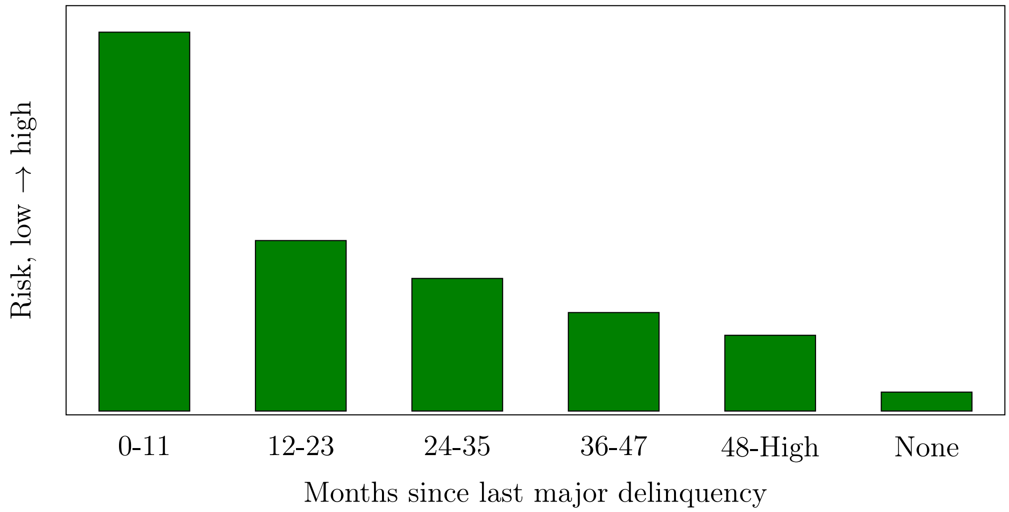


Figure 4.20: Effect of the months since the last major delinquency in FICO .

  - The damage a delinquency does to the credit score increases more than linearly with the number of occasions / missed payments / delinquent trades. The effect of two delinquencies for instance will be much larger than one could assume from doubling the effect of one.
  - It is also worth noting, that even the worst credit scores encountered will have $95^{+}\%$ of payments made. Good customers far exceed 99%.

- **Amounts owed** for 30% on FICO will look at how well you handle credit and whether or not you are deemed overextended. How much owed is too much? A graph on the relative risk of different levels of credit card utilization is given in figure 4.21 (from FICO credit boot camp).
  - amount owed across all accounts
  - amount owed per type of credit
  - how many accounts have balances
  - how much of available credit is used? Focus on revolving accounts such as credit cards
  - how much of an installment loan is owed versus original amount

- **Length of credit history** for 15% at FICO . An illustration of the relative risk of various ages of the oldest revolving credit line is given in figure 4.22 (from FICO credit boot camp).
  - how long have you had your oldest account, newest account and all accounts on average
  - how long have you had specific credit
  - how long has it been since last use

- **Credit Mix** for 10% at FICO looks at all accounts mentioned above. Becomes more relevant the fewer other relevant information your report contains. Looks at

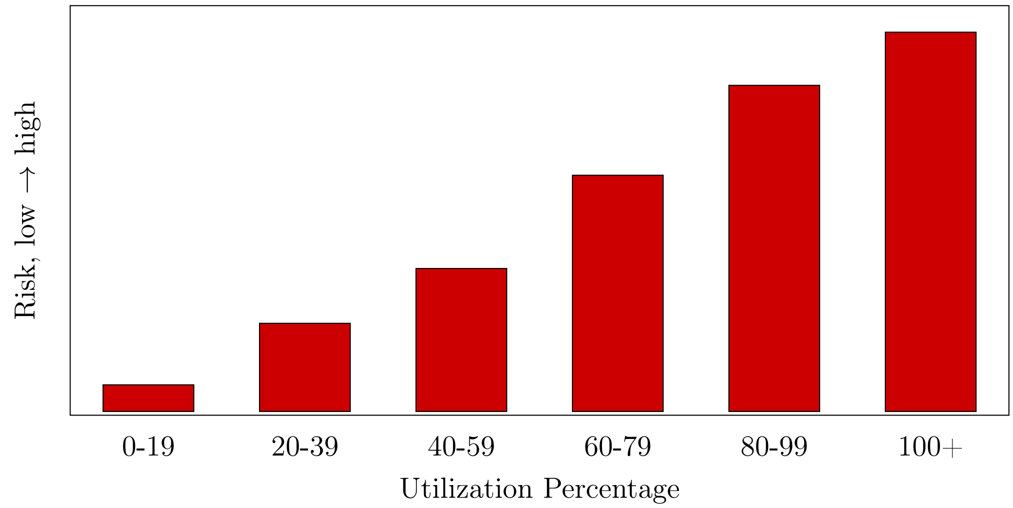


Figure 4.21: Effect of Utilization of available revolving credit in FICO .

what types and how many

- What is the mix of credit products?
- How many bankcard trade lines / revolving credit ?
- Percent of trade lines that are installment loans

New Credit for 10% at FICO estimates risk based on recent credit lines. Soft inquiries such as requesting the reports for information without actually from authorized organizations such as **AVANT** do not damage this score, hard ones do. A visualization of the effect of the number of inquiries on risk for mature as well as thin file consumers according to FICO credit boot camp is given in figure 4.23.

- how many new accounts you have, new accounts also adversely affect the credit history because of the average account age, so fewer in short time are better.
- how many recent inquiries have you had has a small effect if any, a large number of e.g. auto loans or mortgages are not treated as individual requests, but lumped into one with minimal impact on score
- only inquiries from the last twelve months count
- inquiries within a short time span (e.g. 45 days) that indicate pricing comparisons for mortgages, car and student loans will be consolidated into a single one in the summary variables for newer scores if older than 30 days and will be ignored if they posted in last 30 days
- only customer-initiated inquiries for credit posted in the last 12 months will be assessed; Promotional, Account Review, Consumer Disclosure, Insurance and Employment Inquiries will not be considered
- how long has it been since you opened your last account
- is your current history better and does it overcome your past issues

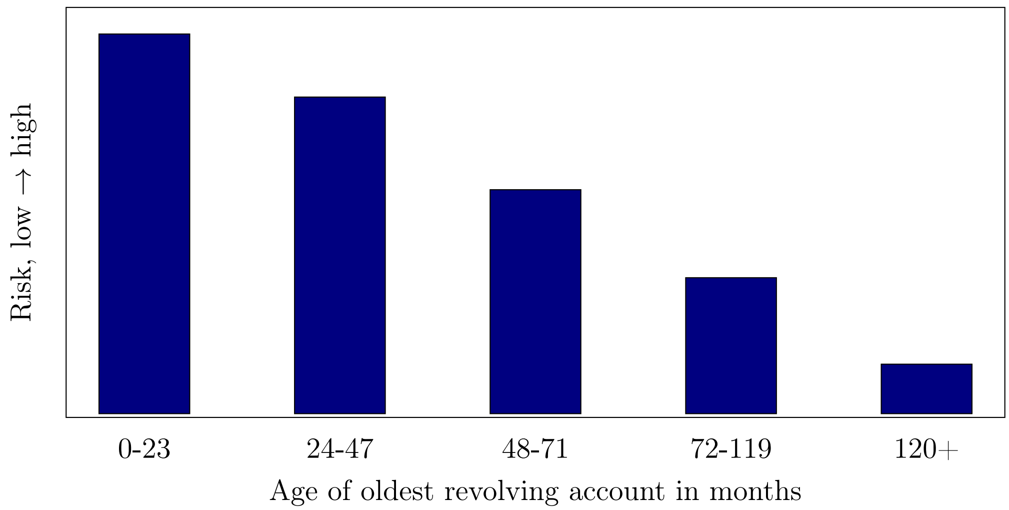


Figure 4.22: Effect of the age of the oldest revolving account in FICO .

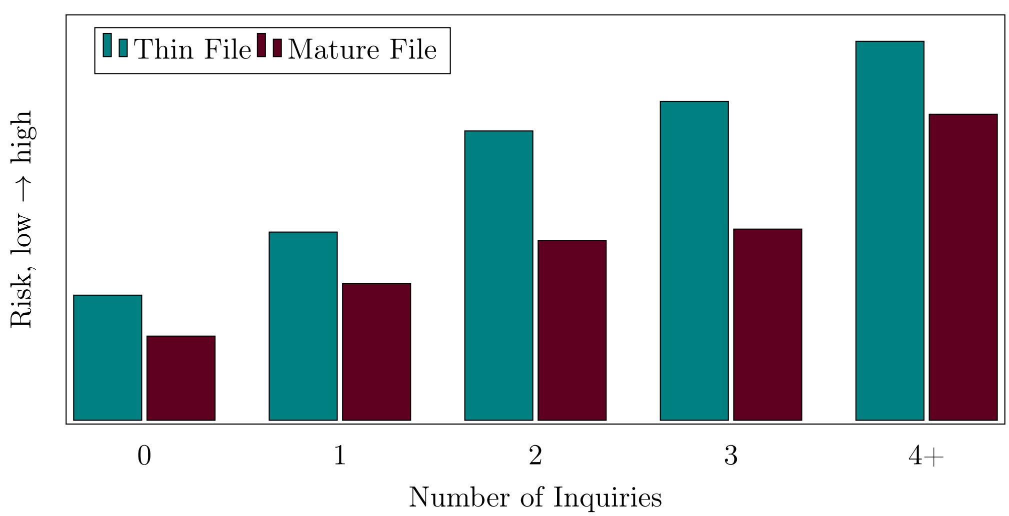


Figure 4.23: Effect of Inquiries in FICO .

Table 4.5: The Gini parameter in percent for Vantage score version 3 on various populations. On the expanded addressable universe the Gini is 54.78%.

Industry	Account Management	Originations	Δ
Bankcard	81.33	76.29	5.04
Revolving	80.10	74.88	5.22
Installation	74.15	67.50	6.65
Auto	75.50	68.14	7.36
Retail	78.85	72.61	6.24
Real estate	77.51	69.65	7.86
Department store	79.73	75.55	4.18
Credit union	75.75	61.46	14.29
Student loan	72.51	66.19	6.32
Overall	79.49	73.47	6.02

4.3.3 What makes a Vantage Score?

do some expansion on vantage score, vantage score history, the different weighings of vantage 1,2,3,4...

The Vantage Score has been introduced in 2006 by the three bureaus to harmonize scoring across themselves[8] as well as break the dominance of FICO. The original score range is 500-1000 for generations 1 and 2. Starting at generation 3 the range has been adopted to more closely match the common FICO range of 300-850. The Effect of the different ranges can be seen in figure 4.24. It would appear as if the scores are vastly different, however, it is just a score calibration as can be concluded form the analysis by quantile of population. this is demonstrated for account originations in figure 4.25.

The fourth Generation of vantage scores is expected to become publicly available in Q4 of 2017.

Which credit factors go into a Vantage ?

- **Payment History:** Repayment behavior (satisfactory, delinquency, derogatory)
- **Utilization:**
  - Percentage of credit limit used
  - Proportion of credit amount used/owed on accounts
- **Balances:** Total amount of recently reported balances (current and delinquent)
- **Depth of Credit:**
  - Age and type of credit
  - Length of credit history and types of credit
- **Recent Credit:** Number of recently opened accounts and credit inquiries
- **Available credit:** Amount of credit available

The details of what goes into each segment is mirroring that of FICO as discussed above.

[8] Introduced Universal Scoring as each bureau had a slightly nuanced way of computing FICO scores owing to the subtle differences on how data is reported, stored and summary variables are built.

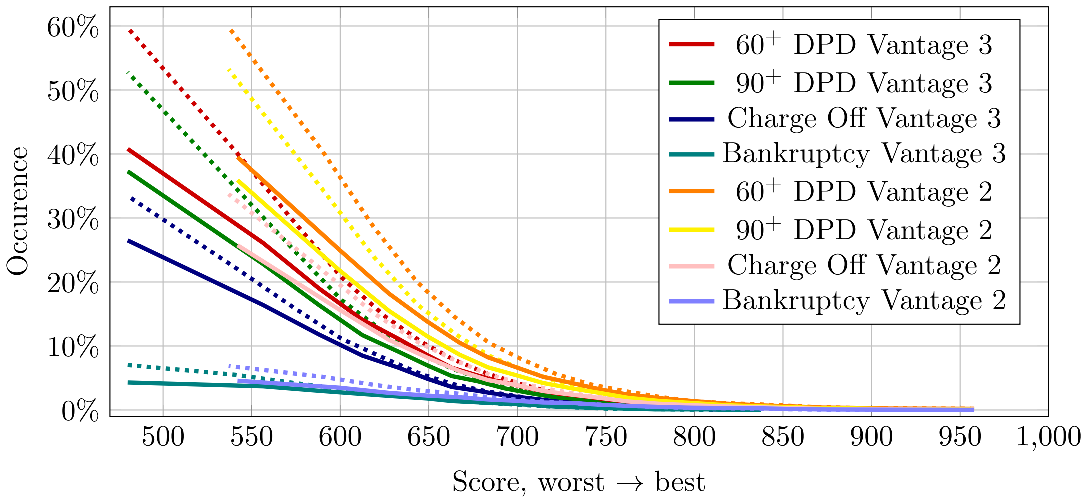


Figure 4.24: Different targets by Vantage Score generation. Solid lines are for originations, dotted lines for account management. Vantage 2 and 3 feature different scales (500-1000 vs. 300-850) which can lead to misinterpretations. E.g. it appears that the Vantage 2 originations charge off curve coincides with the Vantage 3 90^{+}DPD curve and one might assume that version 3 is vastly better than version 2 which is not the case. See figure 4.25.

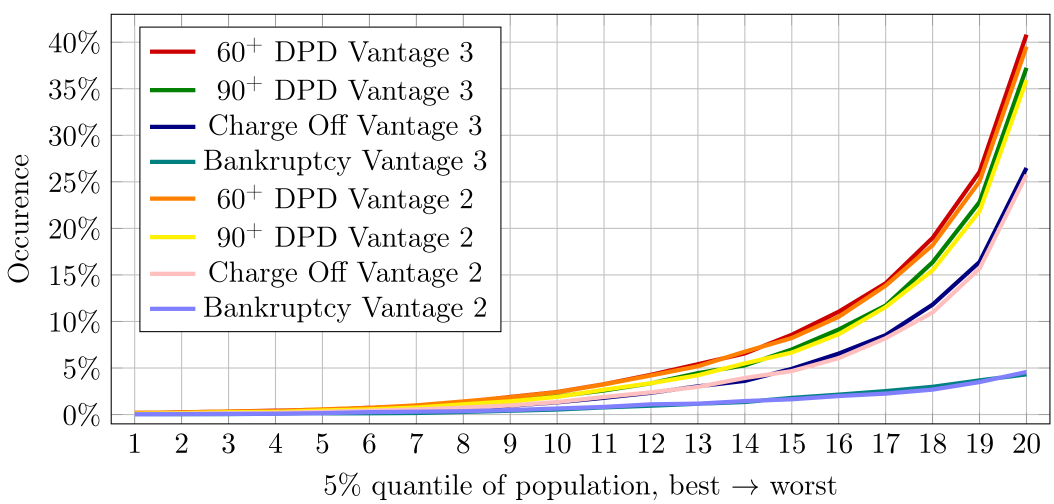


Figure 4.25: Originations by quantile of population for different generation Vantage Scores. It can be clearly seen that vantage scores 2 and 3 provide a very similar ranking of the population even though the absolute score ranges differ as seen in figure 4.24.

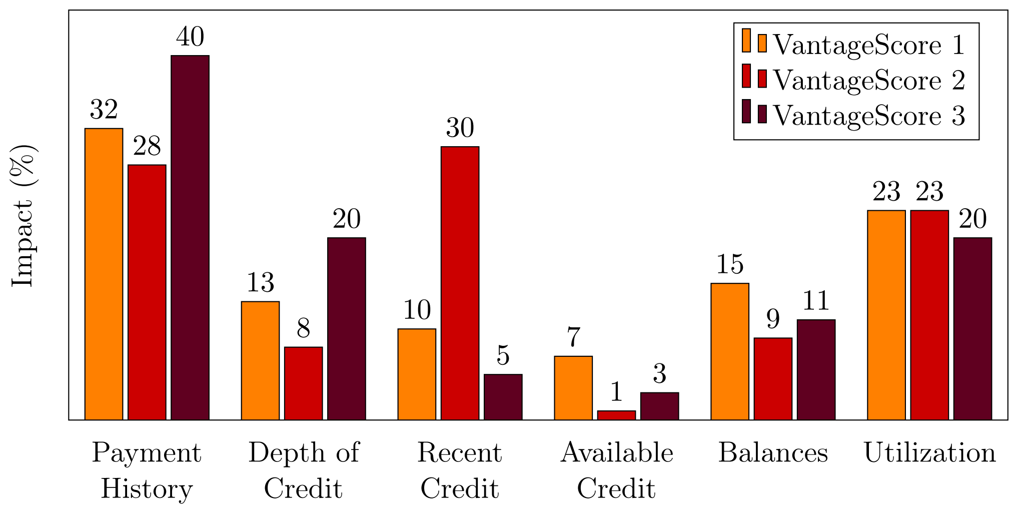


Figure 4.26: Impact Factors of Vantage Score. Weight of categories in Vantage scores. For version 2, recent history has a much higher weight than in FICO which does not carry to later iterations such as Vantage score 3. The recent credit and search history was the dominant predictor during the time of design of Vantage 2 which happened to coincide with the great recession of 2008.

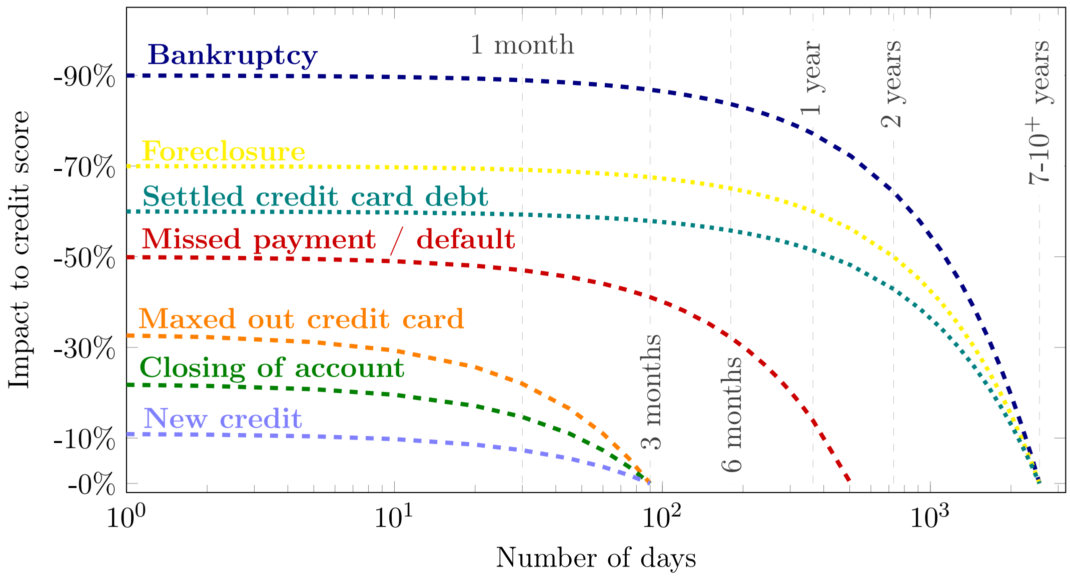


Figure 4.27: Impact of negative credit report items on your credit score. Data taken from Vantage (dashed) and developed from FICO (dotted). The recovery is basically a linear drop over a fixed amount of time, so we could correct for that very easily and distinguish a recently maxed out credit card (very relevant) from a five year old bankruptcy, foreclosure or settlement (doesn't really matter to us).

4.3.4 How and for how long does adverse action affect your credit score?

Adverse action affects your credit score by lowering it by a specific amount over a set period of time. According to Experian , bankruptcies have the worst and most profound effect going for 90% impact and 7 to 10 years. Open collection accounts are second, however, even closed ones stay on your report for up to seven years. Tax liens stay for 7 years, unpaid ones **indefinitely**. Settled debt has a huge impact for up to seven years.

A defaulted payment will impact your credit score around 50% and lose its effect over a 18 month period. Maxing out a credit card, closing accounts, and new credit only affect scores for up to three months. These minor entries stay on your report for two years, positive data indefinitely. Figure 4.27 helps visualizing the effect of adverse items.

What does this mean for a company focused on short term lending? Our models should be made sensitive to score changes based on adverse action that is recent as opposed to those that lay long in the past and do not influence current customer behavior. A recently taken credit is much more relevant to us than a foreclosure 5 years ago even though the impact on the most frequently used FICO score is roughly the same.

Figure 4.28 corroborates the near linear dependence with time of several negative public records. Graphs are compiled from our database and cleaned up to look at singular effective negative entries.

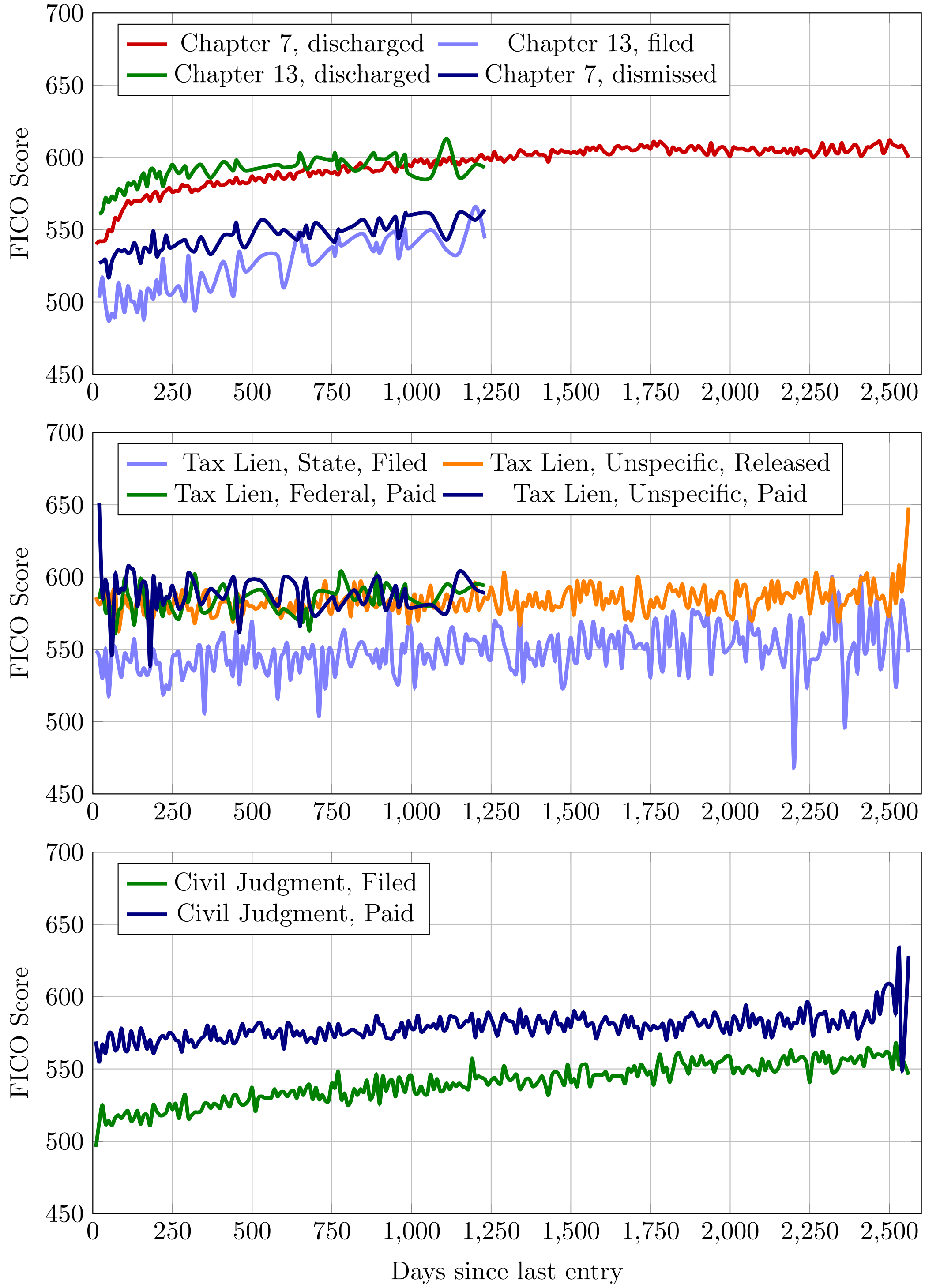


Figure 4.28: Average FICO score in dependence of the time of the last record from the Big Book of Boo Boos (public records) taken from our database on January 30th 2015. Only entries with more than 3,000 records have been used, and only when said entry was the only entry in the report so as to isolate the effect.

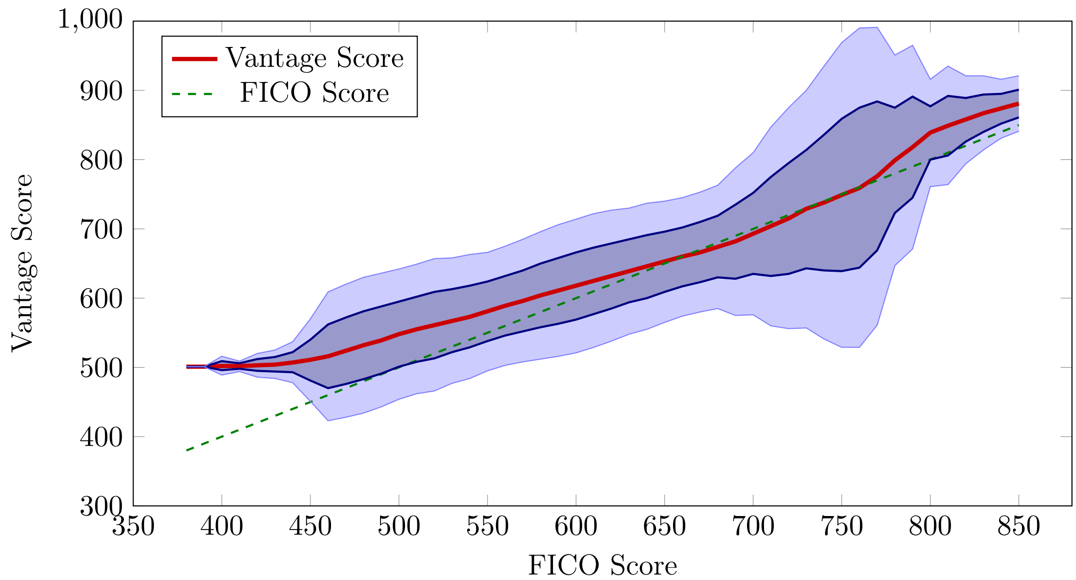


Figure 4.29: Average Vantage score version 2 plotted vs. FICO score version 08 according to internal data. Shaded areas are one or two standard deviations respectively. Dashed line shows FICO score. Apparently for low scores the discrepancy increases, however, in the most relevant areas they are pretty close. At scores around 750 the deviations are substantial in both directions even though the sample size is sufficiently large.

4.3.5 Why would anyone not have a credit report or a credit score?

Experian estimates that about 64 million consumers in the US are unscorable due to insufficient data. Often these consumers are immigrants, recent college grads or have not had an active credit account in the last six months. A large portion (estimates vary between 30 and 35 million) of this group may become scorable with alternative data and scores other than FICO such as the Vantage . About a third of these consumers is expected to be most desirable for lending at the equivalent of FICO score of 600 and above. A good read is found at CNN Money.

The CFPB puts the number of scorable American consumers at 80+% or 188.6M. An additional 19.4M (8.3%) have unscorable credit records, split almost evenly between insufficient for scoring and stale unscored. The remaining 11% (26M) are credit invisible (no data in report). FICO puts those numbers at 170-200M scorables, 28M with reports lacking the minimum scoring criteria (sparse or inactive credit files) and 25M with no records.

All credit scores in circulation have minimum scoring criteria that need to be met in order to generate a score out of a credit report.

Minimum scoring criteria and alternative data use - FICO

According to FICO , three minimum requirements must be met to get a FICO score:

- At least one account has been open for at least six months,
- At least one undisputed account has been reported in the last six months,
- No indication of deceased on the report.

In particular, a report with old trades or devoid of trades, containing only public information cannot be scored.

Experian states that credit reports requires to have at least one credit account open or opened with a reporting lender in the time-frame before the information goes stale and is deleted from file.

The reasoning why a minimum scoring criterion is required can be found in FICO's white paper on Unscorables. Long story short, scores based on stale information lose predictive power, while maintaining a ranking. It is not, however, consistent with the odds of regular FICO scores. The points to double odds can easily increase by 50% for staleness over 24 months vs. currently active consumers, say from 30 to 45 points. This number increases to 60 if the latest activity on the report is older than 48 months. Overall, this lack of accuracy may result in higher than desired interest, decline or acceptance rates, generally, unsafe risks in lending.

FICO gives the following segments of unscorable consumers with a credit file:

- Involuntary Credit Inactive 39% (≈11M consumers)
- Voluntary Credit Inactive 25% (≈7M Consumers)
- Inquiries only or credit for less than six months 10% (≈2.8M consumers)
- Collections or adverse public records only 26% (≈7M consumers)

The same reference also gives a comparison number for the Gini coefficient of the FICO score on scorable consumers with new history vs. first time borrowers. The Gini coefficient is rather low for both populations at ≈47% for the scorables, but in the order of 5% for the new to credit population. Alternative data or data that is not contained in a conventional credit report may help solve this problem. The data selection for use as alternative in the scoring process is guided along the lines of

Regulatory Compliance the data source must be compliant with all regulations pertaining to credit even if its original sourcing was not regulated as such

Depth of Information Data sources that are deeper and contain greater detail provide greater value

Coverage Scope and Consistency: A stable database covering a broad percentage of consumers can be favorable.

Accuracy How reliable is the data? How is it reported? Is it self-reported? Are there any verification processes in place?

Predictiveness The data should predict future consumer repayment behavior

Orthogonality Useful data sources should be supplemental or complementary to what is captured by data sources already used.

Alternative data with most promise are

- **Telco and Utility Data**
  - payment information with similar characteristics to traditional credit file data
  - broad and growing coverage ≈186M consumers (2015)
  - Includes Mobile , Landline and Cable payment history

  - Data Source: Equifax and National Consumer Telecom & Utilities Exchange
- **Public Records and Property Data**
  - Most inclusive of data sets investigated
  - Includes sources from tax and government bodies, USPS (e.g. address changes)
  - Data Source: LexisNexis

Using this data on the new to credit population pushed the Gini coefficient up to ≈40% which is quite a bit lower still than established scorable customer segments but a lot higher than without the alternative data.

When using alternative data in a FICO score, a large number of consumers becomes scorable. FICO breaks it out as follows:

- 45% of Involuntary Credit Inactive now scorable (4.9M)
- 43% of Voluntary Credit Inactive now scorable (3M)
- 76% of Inquiries only or credit for less than six months now scorable (2.1M)
- 49% of Collections or adverse public records only now scorable (3.5M)
- and finally: 54% of consumers without credit history now scorable (14M)

All in, this resembles 28 million or 53% of previously unscorable consumers that are now scorable.

The score distribution of previously unscorable consumers is on a classic FICO scale heavily lopsided towards lower scores, however, about a third has credit scores higher than 620 and ≈10% score higher than 700. Of consumers with alternative FICO score larger than 620 78% had a higher than 620 FICO 9 score 2 years after the credit application with 49% exceeding 700.

Minimum scoring criteria and alternative data use - Vantage

Vantage Scores segment a little differently and claim to be able to score up to 98% of all consumers with a credit file. Unscorable populations segments are for instance (see Vantage Score White Paper

New to market All trades are less than six months old

Infrequent user No trade has been updated within a six month window.

Rare credit user No Activity on their file in the last 24 months. This is the largest group and contains mostly consumers who have not had a need for credit in the last few years, e.g. house is paid off, car is paid off, kids' college is paid and earn reasonably well.

No trades A sub-prime population with closed trades, public records and collections information only.

Expand, make graphs, credit invisibles and philosophies, see webinars

4.3.6 Additional interesting learnings from the data

At this point I shall work out some interesting numbers from the data at hand. Using the Transunion credit reports that **AVANT** has stored in its database, we can compile a few interesting facts.

1. While by law there cannot be any use of the age of an applicant in a credit decision or a credit score, a trend if not correlation emerges when analyzing both sets of data as shown in fig. 4.30. Proxy parameters such as account age may be used for that.

   It is important to know that the age of the applicant was computed from the entry in people.db which is possibly severely biased as we do not create that record if the customer is not loan eligible. However, while the numbers may be different when unbiasing the sample, the trend should come up regardless.

2. Figure 4.30 also shows an interesting statistic on our customers, namely count of reports by age. Clearly we are more attractive to customers of younger age, with an additional spike at the midlife crisis age of ≈ 44 years.

3. The large standard deviation on salaries by age may indicate that a single Gaussian is not a good representation of the data.

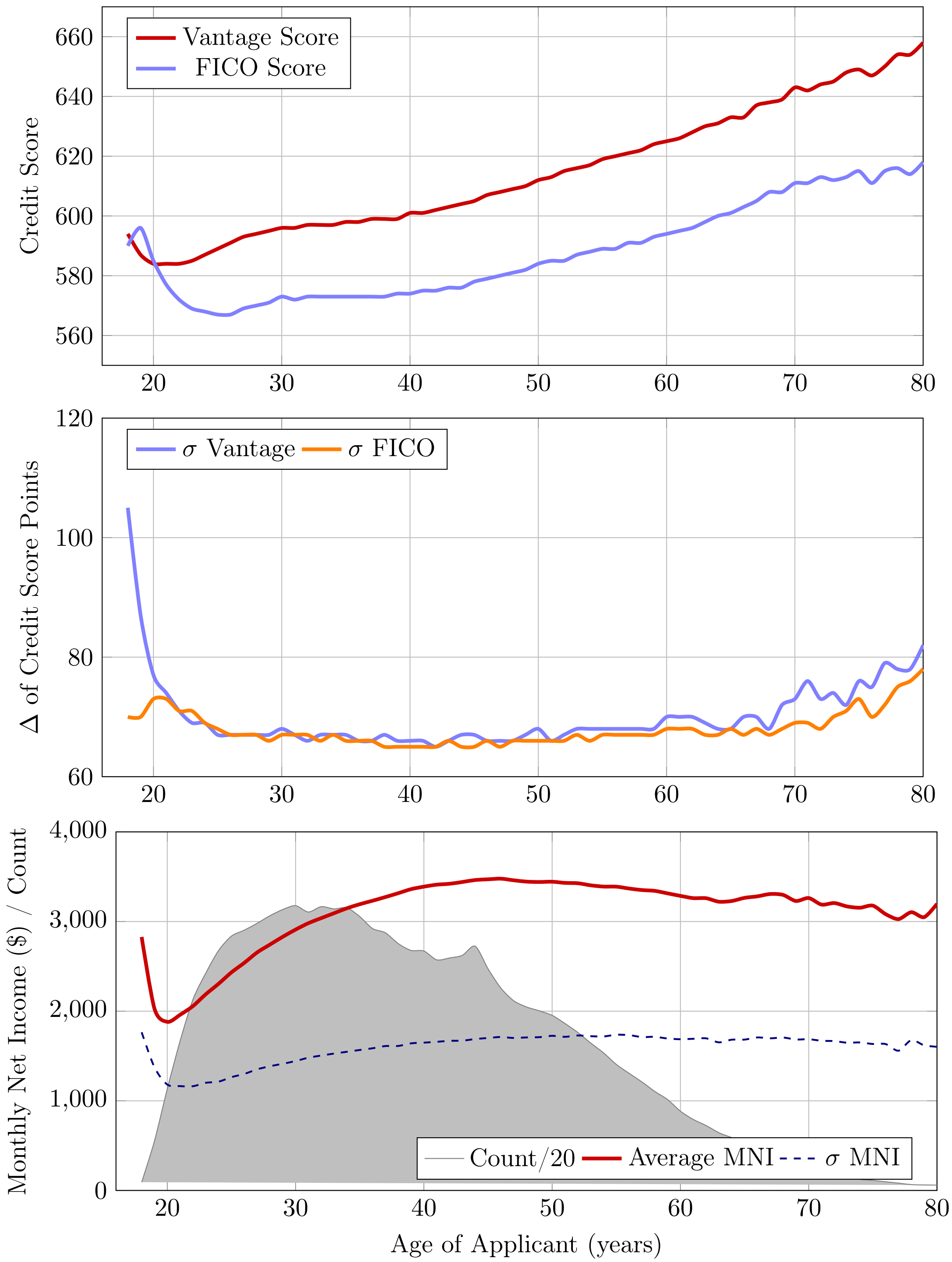


Figure 4.30: Development of credit scores and income with age of applicant. Only ages with more than 1,000 data points are included of this possibly biased subset (see text). **We are legally not allowed to include age as a parameter in a credit decision.** However, proxies such as *age of an account* may be helpful since a clear trend is evident.

4.4 Reference

Reference & Literature:

All graphs and probabilities, odds charts and definitions in this section have been created using data published by TransUnion and the Fair Isaac Corporation.

In particular, the following publications have been used:

- VantageScore 3.0 Validation Summary (TransUnion)
- FICO Score 8 based on Transunion Data - Odds Charts (TransUnion)
- Introduction to Scorecard for FICO Model Builder (the Fair Isaac Corporation)
- FICO Score Fundamentals, Julia Wooding, Fair Isaacs Corporation Credit Boot Camp, Webinar, 2014
- To Score or Not to Score?, FICO White Paper, Number 70, September 2013
- Universe Expansion: Is the Way You Score Customers State of the Art or State of Denial?, Vantage Score White Paper, March 2014
- Data Point: Credit Invisibles, K. Brevoort, P. Grimm, M. Kambara, CFPB White Paper, May 2015
- Inside the FICO Score and Alternative Credit Data, FICO webinar, July 22nd 2015

Chapter 5

Models

data sciency bla bla

5.1 Introduction or When do we need a model?

many parameters, no easy to discern business rules, complex relationships, reasonable effort for estimated gain.

Figure 5.1: High Level How Avant Works (Executive Summary)

5.2 Model Lifecycle

This section outlines the high level steps that are involved in getting a model from the wish list into production. Items written in `typeface` are subject to debate between teams and need to be agreed upon prior to usage of model and handled with care and caution.

1. **Business** drives the model creation / iterations. **Business Analytics** (**BA**) will oversee the details of the process.
   (a) Define the goal and character of the model. Is it a refit to an existing model or something new? Is it a major change to the character of the model such as changing the purpose or is it a refit with just more data added?
   (b) Define what data sources we want to use
   (c) Have a good grasp on which `indicator` (or group of `indicators`) we want to use directly or as a proxy of overall performance.
   (d) Have proof that what we want adds to the existing performance. Preferably all new features have been tested in a simulated non-production tool such as data-robot. Ideally we can also show that with application of simple rules we cannot improve as much.
   (e) Ensure that desired capability is not duplicated by a different `indicator` or model. Provide `indicator` comparison or percentile plots of actual data versus mean `dependent variable`[1].
   (f) Ideas are typically generated by availability of data, business goals, new vendors, competition pricing, best industry practice, our list of not-so-low-hanging-fruit, anything we stumbled upon when looking at the data or any insights, cross product / geography communication[2].
   (g) Update shared documents such as Hackpad to reflect the process.
2. Planning phase with **BA**, **Modeling team** and **Engineering**.
   (a) Agree on timeframe, variable inclusions, availability of data
   (b) Scope out **Engineering** needs. This step puts all else on hold, if not feasible, model will proceed without all desired variables at this time. Prototype models can be generated outside the **Engineering** environment though.
   (c) Integrate project into epics on **Modeling team** pivotals.
3. **Modeling team** pulls data
4. **Modeling team** shares `indicator` data and new variables pulled with **BA**. An explicit sign-off is required of both teams.
5. **BA** and **Modeling team** together validate the input data and usable `indicators`
   (a) **BA** team verifies that data used for modeling is in line with expected data from independent querying.

[1] An example would be a *first installment 60 days past due* `indicator` is a subset of of the `indicator` *not current 7 days after 3rd installment* as used in UK model. It showed no additional benefit without additional data sources.

[2] If something is interesting in the UK why would it not be interesting int he US? Other than legal constraints, of course

(b) This ensures that the production sources of model variables match the sources used in **BA** .

(c) The **Modeling team** use sources pulling data into the modeling environment, that are not necessarily available on the production site in a similar fashion (e.g. Ruby vs. R).

(d) For the same reason, it is essential that the indicators match between **BA** and **Modeling team** data.

(e) Popular / frequent problems include but are not limited to:

- `customer applications` data identification as **Business Analytics** follows a `lead` in a straight mapping to issued `loans` , whereas modeling goes from issued `loans` backwards which can include several decisions per `customer application` , several `customer applications` per `loan` , several `leads` per `customer application` and several `reports` at different times in the process. **A picture tells a thousand stories.**
- mismatch of cached and production data,
- mismatching joins between data available for issued `loans` vs. data that was used to issue the respective `loans` ,
- Locale pollution on shared service modules,
- ...

6. **Modeling team** models (predicts) with appropriate methodology and determines `indicators` that can be predicted efficiently. The result at this point is a force ranking of `customers` based on the information used in the model to predict the `dependent variable` / `indicator` . This we refer to as raw scores[3].

7. A more extensive discussion of the modeling part can be found in section 5.3.

8. **BA** and **Modeling team** choose `indicator` best suited to business goal (highest correlation to desired outcome) as well as most efficiently predictable[4].

9. **Modeling team** creates a `validation set` from most recent issuance excluding the `loans` that would not pass the current business rules as those will be declined automatically and not by the model[5].

10. **Modeling team** will generate a score formula translating the rank order into probability of `dependent variable` . This we refer to as transformed scores. A good

[3]Explain bla bla, difference to transformed scores.

[4]Example: say the business goal is to reduce charge-offs in the first six installments after issuance. This requires seven months of time to pass after issuance of `loan` to reach six installments plus an additional 120 days past the last installment to allow for collections and recovery effects. This is almost a year of time, therefore significantly reducing the available data set as well as increasing the time required for monitoring quantitative model performance. We would want a shorter term `indicator` with a high correlation (prediction) of the long term behavior such and an `indicator` that has great modeling properties such as high separation of bads from goods in a smooth fashion **graphics**

[5]We are looking to improve the decision making on those `loans` we cannot decline / target with simple business rules

transformed score matches the actual distribution of the `dependent variable` in the range of interest[6]. **Some text on straightening the line**

11. **Modeling team** will generate a model card including the following items:
    (a) data sets used
    (b) reference to model objects
    (c) findings on `indicator` predictability for various `indicator` combinations of `indicators` used on `training set` vs. `indicators` used on `validation set` [7].
    (d) most important variable lists
    (e) partial dependence plots for most important variables (or more)
    (f) ...
12. **BA** reads through model card.
    (a) What are the top variables? Are they in line with expectation?
    (b) How do top variables relate to top variables of previous models predicting same outcome or similar target group? Are they the usual suspects or far off?
    (c) How does partial dependence look for variables used by third parties?
13. **BA** will then confirm the `validation set` findings of the **Modeling team** .
    (a) Confirm eligibility of data used in validation. (Eligible performance data, matching `dependent variable` , etc.)
    (b) Confirm score formula.
14. When score formula is confirmed, **Modeling team** retrains model on the entire available data set (`training set` and `validation set` combined) to take advantage of the larger quantity of data.
15. Modeling and **Engineering** teams deploy model for parallel production (model scores but is not used for any decisions yet).
16. **Modeling team** scores recent `leads` , time span covered determined by variability of `lead` volume. Limit scoring to only `leads` not declined by anything else (rules, other models etc.). This step enables us to reduce time to production use of the model.
17. Above two steps will help determine quantitative impact of new model. See section 5.4
18. Get numbers input from all teams involved.
    (a) **Business** and **Product** targets, usually pertaining to volume projections and tier distributions / impacts,

[6] No need to fit over far outliers at scores of 80% when interesting range is below 20%. The best fit for immediate pricing needs should get hte most attention / weight. Cut offs should not be applied too tight.

[7] It is possible and not rare that an `indicator` A used on `training set` will not only predict `validation set` with `indicator` A best, but also predict `validation set` with `indicator` B better than using `indicator` B in `training set` .

(b) **Finance** allowances from your favorite **FP&A** team, for target profits and allowances of bads,

(c) **Capital Markets** and **Compliance** for any kind of distribution concessions they made in order to secure licensing or financing, limits here are usually centered around the issuance composition[8],

(d) ...

19. **BA** will facilitate necessary product design changes if any.

20. **BA** will involve **Engineering** and **Developer teams** to activate model.

21. Release will be watched immediately for all desired outcomes / compositions by **BA** , **Modeling team** , **Developer teams** , ...

22. Performance will be monitored for a long time by **BA** **see extra section on this, focus on ranking and rough numbers as as soon as available**

[8]Example for **Capital Markets** request: limit exposure in tiers with internal rates higher than 40% to less than 30% of money issued or x £ whatever is less. **Compliance** may ask to find some public disclosures in need of alignment such as more than 50% of issued loans by count have to have internal rates under 40%. Etc. If those two group's asks align with the others, congratulations, usually they do not and need to be balanced.

5.3 Overview on what models do and modeling techniques

force rank customers based on information at hand to predict their likelihood to do y

5.3.1 logistic regression

5.3.2 tree models

include table for comparison of methods, regarding data that can be fit, pro con, timeframe, future bla bla, all the other types, income model, autoapproval, marketing, ...

5.3.3 How to tell what is a good model

5.4 Model Validation

5.4.1 Common Sense, Conventional Wisdom and Sanity Checks

5.4.2 How to mathematically validate a model

There are several fundamentally different uses of the models provided

Bla bla on pricing models.

Typical tasks for continuous variable models are for example:

- price `customers` by risk, e.g. tier pricing by probability of delinquency in the first y months of taking out a `loan`
- price `customers` by amount of money expected to be earned in timeframe z
- models that predict response of `customers` to marketing efforts such a response models
- ...

Cutoff models are used to pass or fail a set of `customers` based on a threshold score. Typically cutoff models are used for a `dependent variable` that occurs about an order of magnitude less often than the `dependent variable` used to price customers, the effect of which is clearly defined and always bad[9].

Typical tasks for cutoff models include:

- cut x_1 percent of overall count / percentage of bad property y
- cut x_2 percent of y to maximize overall profit
- cut x_3 percent of y to maximize the profit per loan
- drop x_4 percent of y from population qualifying for process z (e.g. auto approvals, fraud checks etc.)
- ...

Designing a model cutoff is to find the optimum tradeoff between bad `customers` cut by the model and loss of good customer volume. This ratio should exceed 1, ideally a larger number. A 2:1 gain on all configurations at least is desirable, higher is better. We also want to be conscious of our volume implications as we can very easily cut all bads if we just cut all loan issuance. This is - in the words of **AL GOLDSTEIN** - *a money losing proposition* and will therefore never be signed off on. Ideal cutoffs target low single digit percentages of volume for high percentages of historic bads.

[9]Example, it is acceptable for our `customers` to become delinquent with 10% probability in the first three installments since we can put collection efforts behind that and get most of our money back. It is not acceptable for us to issue `customers` with fraudulent intent (no intent to pay, also referred to as *straight to charge-off*) at a rate of 1.5%. Those would be full losses and detrimental to our overall performance. In this case you would price delinquencies to `customers` in a tiered pricing system, giving those with lesser likelihood of delinquency better rates and terms than the ones with higher predicted likelihood and use a cutoff model to target say 75% of the fraud cases to get to a 0.4% straight to charge off rate (which would be a great number to achieve in the financial industry).

5.4.3 Required input data

The data science team will have provided a new model to be assessed. We need:

- Regardless of model type we need the `validation set` including:
  - `loan` id,
  - raw and transformed model score of model to be assessed,
  - `dependent variable`.
  - If and when we look at the entire data set, we need a marker for `loans` that were used in `training set` or `validation set` respectively
  - It is usually a good idea to include all and any variable provided by third parties or customers that are aiming at a prediction of the `dependent variable`[10].
- the following are required only for cutoff / pricing analysis and not for score validation:
  - scored set of data in parallel production with `lead` id and model scores
  - average `loan` amount issued A_L, usually historic, but can be from `leads` scored
  - estimated or expected profit per `loan` P_L, ask the **FP&A** team
  - a side computation that will cancel out in the end is the estimated number of `loans` issued in the given timeframe of scoring N_L
  - and optionally model score of models applied prior to running this new model
- for quantitative prediction it is advisable to include in the `leads` set information about
  - import decision
  - previous model score if we are replacing a model,
  - `loan` amount
  - issued flag to assess conversion behavior
  - additionally for pricing design and conversion analysis it is good idea to see if the customer actually came to our site

[10]Examples would be FICO / GAUGE / Risk Navigator / Fraud / indebtedness / VANTAGE scores

5.4.4 Validation

1. Mandatory checks:
   - What is the size of the data set?
   - How does it relate to the size of available data? Make sure we have the full `validation set` and ONLY the `validation set`[11].
   - The `validation set` needs to be *limited* to current issuance[12].
   - Pull the `validation set` `indicator` (`dependent variable`) from an independent source (not same team that made the model). Check for matching `indicator`, if `indicators` do not match, find out why. Validate against `indicator` you know to be most accurate. If there is no correlation and no reasonable explanation is found, the model needs to be retrained with the proper `indicator`.
2. Bucket raw scores in deciles (or any other quantiles) **figure out optimum number of quantiles by average `dependent variable` and number of data points, separation (auc) etc.**
3. Determine average model score in bucket
4. Determine `dependent variable` average in bucket
5. Plot average score, average `dependent variable` over deciles
6. Monotonic dependence is DESIRED for pricing models or such, quantitative match is ideal within small deviations
7. For cutoff models there should be a steep increase in average dependent variable and score in the 10th decile compared to all others otherwise the model is not useful as a cutoff model and has to be rejected and / or misappropriated for pricing.
8. Once this is confirmed we APPROVE **score formula**
9. At this point we can use the third party scores for a sanity check and compare prediction by external score bucket. If we are validating a pricing model , this helps with that.

Validation chart, Volume distribution chart for leads and validations

[11] The `training set` will always be perfectly scored, that is the whole idea.

[12] We want the `training set` to include all available data, however for quantitative estimate of predictions, it is essential to not include data that would not pass current underwriting, business or other rules and would have been knocked out regardless of score.

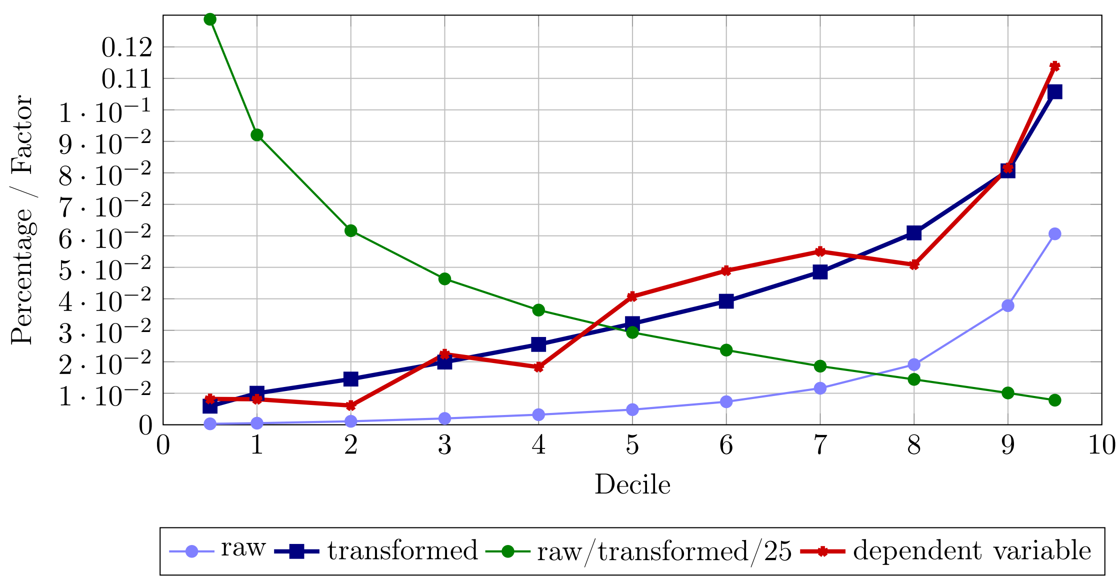


Figure 5.2: Validation plot for the debt management model version 1.1 of February 2016. It shows raw and transformed scores, the relation between the two and the average dependent variable.

5.4.5 Cutoff determination

1. Cutoffs are determined usually by approved score formula values rather than percentiles of rank ordering even though that yields satisfying results as well. It is easier to make a business assessment of impact on a score that is transformed into a probability.

2. Bucket transformed scores into reasonable size bits. It is helpful to start with determining the distribution of `dependent variable` by score bucket of the `validation set` **see figure**.

3. For each bucket i determined by a score cutoff c_i, compute

   (a) Of the `validation set` , how many positives of the `dependent variable` have model scores ABOVE the bucket threshold divided by the number of positives in the entire set (higher is better)? Let's call this ratio d_i

   (b) We can use d_i directly to determine the cutoff c_i for a problem determining x_1 by setting $d_i = x_1$ and use the corresponding c_i to guarantee we were to cut x_1 of the historic bad population.

   (c) Of scored `leads` , how many scored `leads` - that pass other models applied - are ABOVE the bucket threshold aka how much volume were we to cut (lower is better) of total `leads` scored volume V? Call it v_i

   (d) Using a high value of the ratio d_i/v_i is prudent to ensure not to cut too much volume of good. Let us venture a little deeper into that.

   (e) Of scored `leads` what is the average score of the scores ABOVE the threshold c_i? Call it a_i

   (f) Apparently we would lose profit by cutting the goods easily computed as

   $$N_L \times v_i \times P_L, \tag{5.1}$$

   (g) while we would prevent loss of

   $$N_L \times v_i \times a_i \times A_L. \tag{5.2}$$

   In this case it is appropriately wise to dream and use a profit that is higher than the current profit estimates to stay conservative. [13]

   (h) It may also be prudent to use c_i instead of a_i in order to estimate conservatively and know that there is some air cover provided by the score's average.

   (i) the interesting ratio is the amount of money saved divided by the profit lost k_i as described as

   $$k_i = \frac{v_i \times N_L \times a_i \times A_L}{v_i \times N_L \times P_L} = \frac{a_i \times A_L}{P_L} \tag{5.3}$$

[13] A point of concern here, while this straight applies to cutoff models targeting a first installment property, we may actually have some earnings from previous installments in case the `dependent variable` is a certain property in a reference time frame exceeding the first installment as for example the `indicator` *submitting debt management paperwork in the first six installments.* A careful analysis of this would lead to a cutoff value that is **higher** depending on term and time frame. **Elaborate!**

(j) Obviously k_i needs to be larger than 1 to guarantee a gain by using cutoff c_i. In other words, the average expected share of bads above cutoff a_i must be larger than P_L/A_L, our relative profit P_r per `loan` desired. We should therefore choose c_i accordingly.

(k) $k_i > 1$ also means that P_L increases for the rest of the loans $N_L \times (1 - v_i)$ by the amount G of at least

$$G = \frac{N_L \times v_i \times a_i \times A_L - N_L \times v_i \times P_L}{N_L \times (1 - v_i)} \geq \frac{v_i}{1 - v_i} \times (c_i \times A_L - P_L) \quad (5.4)$$

This is the value we would want to maximize. Of course, we can express G as a function of profit P_L:

$$G_P = \frac{G}{P_L} = \frac{N_L \times v_i \times a_i \times A_L - N_L \times v_i \times P_L}{P_L \times N_L \times (1 - v_i)} \geq \frac{v_i}{1 - v_i} \times (c_i \times \frac{A_L}{P_L} - 1) \quad (5.5)$$

Very obviously, our relative profit $P_r = P_L/A_L$ is a parameter in this value. And we stand to make money per loan when $c_i \times A_L/P_L > 1$, or, in other words, if we chose c_i for cutoff to be larger than P_r.

This means, that a safe bet for cutoff is always above the target internal rate of return given by **Finance** and **Business** . On the flipside, it is probably a good idea to look at weighted average internal interest rate of current issuance and choose the cutoff BELOW the amount of interest paid divided by principal per year for that rate with a comfortable cushion[14].

Setting $c_i \geq P_r$ does only guarantee a gain due to the cut, it does not mean that that is the optimum value while $c_i < P_r$ guarantees a net loss. The average value of a_i for $c_i = P_r$ may be a good first idea for the spot of optimum profit. **check this for facts**

charts for profit, chart for different P_r / IRRs

[14] Example: Average rate is 36% this leads to average per annum of ≈ 22% of interest collected. Give or take marketing, reporting and salaries disregarding cost of capital we should probably not aim for a cutoff above say 15%.

Chapter 6

Questions and Exercises

6.1 Loans

1. Create an Excel or Google Sheets workbook that contains an amortization calculator. For any given loan amount, APR and term, compute

   (a) the monthly payment amount

   (b) the total loan amount,

   (c) the total interest,

   (d) how much of the principal amount the interest is,

   (e) how many Dollars per year is interest,

   (f) how does that compare to the APR in percent

   Use Excel / Sheets built in payment formulas (pmt, ipmt, ppmt) **AND** compute it on your own from scratch. Make sure both produce the same results. If there is a mismatch, understand why it is there. Take mental notes on all the things that can go wrong with this because they will at some point.

2. Show an amortization table for at least 24 months for the above loan calculator, for curiosity I recommend to at least look at mortgage financing over 360 months at low APR for comparison. It is very instructive. Investigate the effects of rounding.

3. In the loan amortization calculator above, add the following features

   (a) Add an extra monthly payment applied to principal

   (b) Add a one time payment of $ x at installment y

   (c) Determine at which installment we reach profitability

   (d) Determine how many loans paid off after full term we require to break even on initial defaults

   (e) Determine how many loans paid off after full term we require to break even on defaults after installment k

4. Investigate the effects of changing the term, APR and amount of loan using the calculator

5. Rank the following customers in order of $ attractivity to business providing the numbers including projected default rates etc.
   (a) Joe, $20k, 48 months, 19%, paid back full
   (b) Tom, $5k, 36 months, 65%, paid back full, two late payments at payment 16 and 3, used courtesy program and paid both after skipping the payments, two additional installments added to the end of his term
   (c) Hank, $5k, 36 months, 65%, late payment on payment 12, used courtesy, got back on track
   (d) Al, $5k, 36 months, 65%, late payment on payment 13, did get the extra fee of $25 on payment 14, then back on track
   (e) Julie, $5k, 36 months, 65%, default at payment 33
   (f) Anna, $7.5k, 36 months, 36%, default after 7 payments
   (g) Peter, $5k, 36 months, 36%, default at payment 33
   (h) Maria, $10k, 24 months, 19%, paid back full after 3 payments
   (i) George, $8k, 12 months, 25%, default first payment
   (j) Steve, FICO 680, Income $2.5k monthly, model score 0.01, asking for a 36 months, $5k loan, wears XXXL T-Shirts with Nerd Prints, Sandals and Socks
   (k) Aaron, FICO 560, Income $3k, model score 0.15, asking for a 36 months, $5k loan, dresses nicely, suit and lacquer shoes
   (l) Paulina, had a previous loan paid back full for 65%, 12 months and $2k, wants to take out twice that at a lower rate
   (m) Dimitry, applied with a valid SSN, has valid bank connections, incomes in the US, SSN has been reported with high Fraud probability score, but he is a really nice guy
   (n) Mary Joe, 21, started her life, no FICO score, income $3k, wants $5k for a down payment on a car
6. Given a monthly net income of Ξ, how much principal could a borrower maximally obtain if he is not to spend more than 43% of his MNI on the loan, given A and n? How much interest would that be in relation to the MNI over the length of the loan?
   (a) Compute your results for a reasonable mortgage rate and term!
   (b) Compute the results for a highest tier customer of AVANT !
   (c) Compute the results for a variable percentage of MNI going into the payments!

6.2 Reports and Scores

1. What does the credit score / model score determine and describe?
2. Which parameters are required to determine APR, maximum loan amount and length of term other than legal constraints?
3. Which parameters are not used? Which are most predictive other than third party scores?

4. What is the difference between FICO , Vantage and AVANT model score?
5. How many funded / rejected customers with foreclosures do we have? At which conditions?
6. Is there an optimum number of trade lines, credit cards, collection amounts for the business / the consumer?
7. What is the correlation between the default rate within the first 12 months of taking out a loan with AVANT and our model score? Is there causality? Is there a correlation? Which one?
8. What peculiarities affect the loan conditions per state and in foreign markets?
9. Which information on a given credit report can we use? Which one can we not use? What else can we get our fingers on? For what?
10. What is the logic behind using a specific type of report from a specific provider when there are more suppliers of reports out there?
11. Why does our model score have so many digits?
12. Who is unscorable and why? For each type of report / score?

6.3 A Tale of Ants

In a land not far away there is a gang of three adolescent ants. As is tradition among youngster ants, each possessed the fanciest smart watch ever made. As they set out to their ritual of passing into adulthood which encompasses a long walk through their realm, the following things happened:

As they go into the forest and pass the famous tree, all of a sudden, a loud noise explodes. **"Ka Woom, Ka Pow, Ka Pah"**, the lead ant's smart watches alarm went off. Of course, a speech is expected in which to state a fundamental truth. So it says *"There are two ants walking behind me."* The other two acknowledge, *"Yeah!"* and *"You are the man! You know stuff!"*.

On they go. Pass by the clearing with the meadow when disruptively they hear **"Ka Pah, Ka Woom, Ka Pow"**, the alarm of the second ant's smart watch went off. So it states: *"There is one ant walking in front of me and one ant walking behind me"*. *"You are wise beyond your years"*, *"Right on"* nod the others.

Continuing the trip they pass by the big rock and the puddle they used to swim in when younger, but they are passing it this time and go back into the woods. Here we go again: **"Ka Pow, Ka Pah, Ka Woom"**. The third ant's time to speak has come. *"There are two ants walking in front of me"*, *"Yup, you are so smart"*, *"Excellent"* nods go, and he continues *"... and there are two ants walking behind me!"*

Quiz: How is that possible? Explain!

Hints:

- This can be solved without academic knowledge, there is no funky physics going on such as mirrors, mirages, reflections, Moebius bands, walking on a curved surface or in a circle, Pirouettes, Somersaults etc.
- Most information above is not specific or required to solve this, there is no additional information in the type of animal, smart watch, woods, forest, time of day etc.
- Just to be clear, this is not an IQ, physics or behavioral but a data (science) question
- Stefan welcomes answers, typically people have it in five minutes or never. There goes the challenge!