SGH

Event History Analysis

Practice class I

1. Introduction to Event History Analyses
2. Introduction to Credit risk
3. Definition to basic terms of EHA in application to CR
4. Database

SGH

Introduction to Event History Analysis

Event History Analysis

Survival analysis is a class of statistical methods for studying the occurence and timing of events.
Origins from biostatistics (studies of deaths)

Where EHA methods can be used:

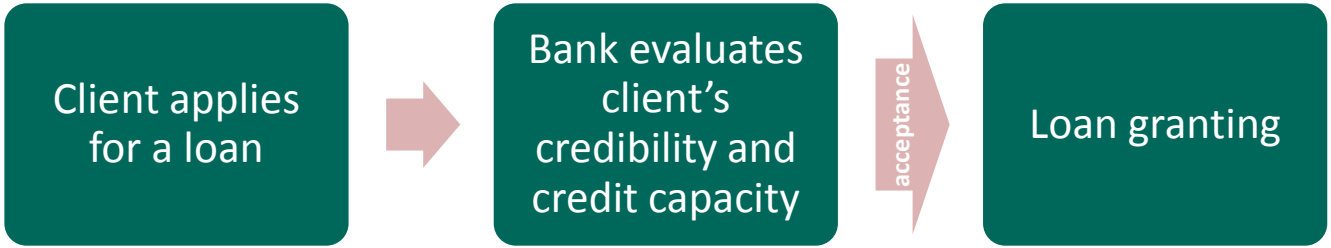
- Demographics – Life Expectancy Tables
- Clinical trials – time to side effects or death
- Insurance – time to insurance claim
- Credit risk – time to default (event of not repaying credit)
- Engineering – time to machine breakdown

Event History Analysis vs. Regression methods

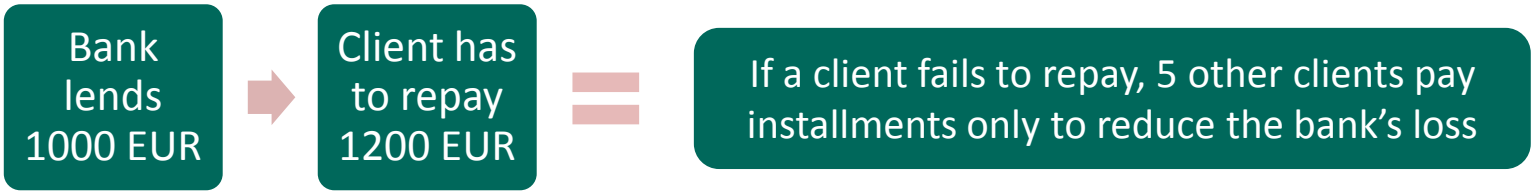
Criterion	Regression	EHA
Target variable	Variety (continuous, discrete etc)	Time to event
Censoring	Full history of subjects admited to the study is needed, especially the target variable must be known	Complex theory about censoring
Outcome	Expected value	Distribution of a target variable.

Introduction to Credit risk

How does the proces of granting a loan look like?



How does a Bank earn money?

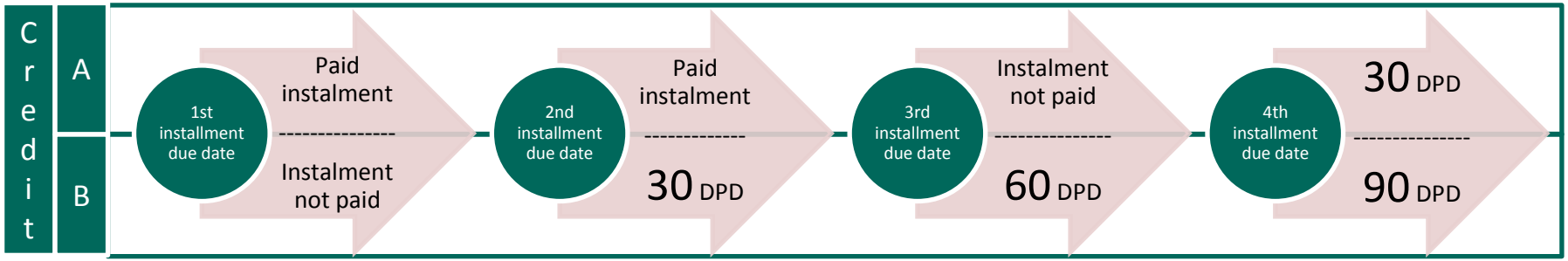


Credit risk is a risk that money granted to a client will not be repaid back due to loss of creditworthiness during the loan period.

How does a bank protect against Credit risk?
→ Especially in Consumer Finance, but not only – they build Credit risk models, whose goal is to estimate the probability of not repaying a credit (=default).

What does it mean „not repaying a credit?”
Definition of default: it is a situation in which a client has a delay in instalment payment exceeding 90 days (known as *days past due*, DPD).

How does the proces of monitoring credit exposures looks like?

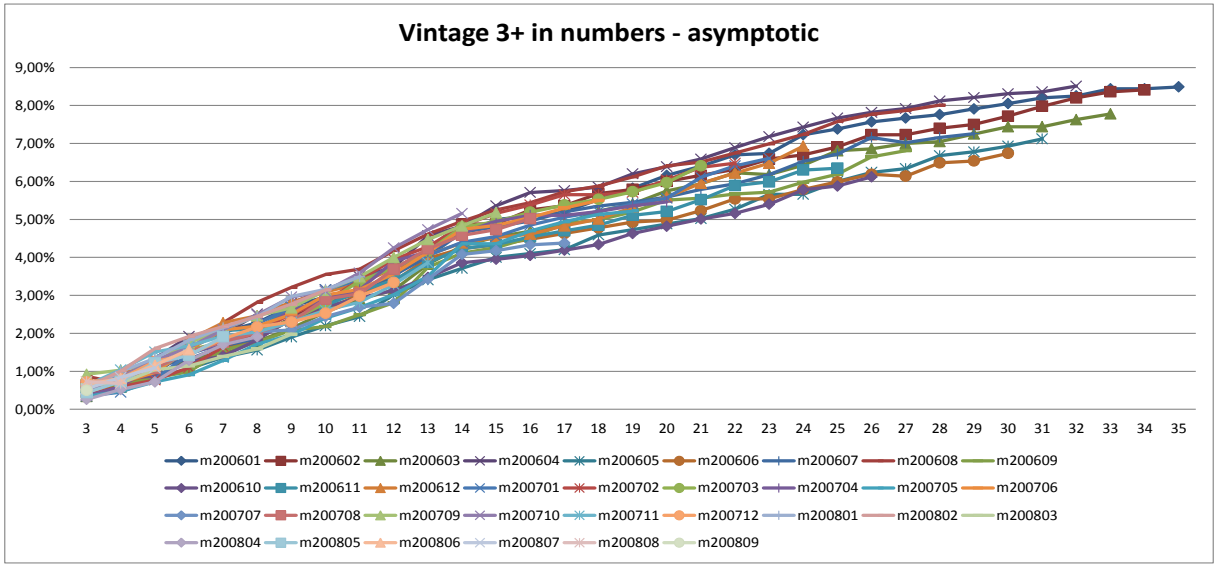


In application model, we model probability that a loan defaults in the first 12 months since acceptance.

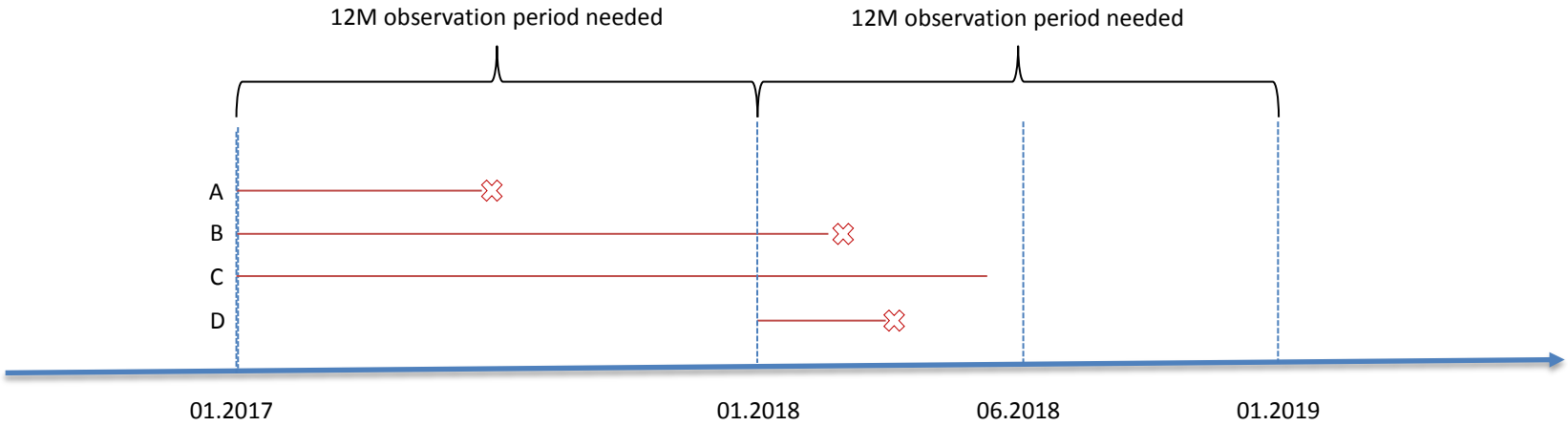
Why is this target variable problematic?

Consumer loans are usually granted for up to 10 years, so an assumption that Credit risk concentrates in first 12M is difficult to pass.

As you can see below the Credit risk during first 12M materializes at 4% whereas lifetime risk is about 8% - banks build models on a risk 4% , but the real risk is about 2 times higher!



Credit risk
Comparison of EHA and classical method



Even when credit policy is being changed, this loan cannot be taken to model building process as the observation period is not completed.

Loan	Classical method	EHA method
A	Default	Default
B	Non-Default	Default
C	Non-Default	Non-Default
D	Not completed observarion period	Default

SGH

Definition to basic terms of EHA in application to CR

Event:

- **Initial** (event-origin) – initiates a career (e.g.: first children birth initiates maternity career)
- **Repetitive** – may occur a few times in a career (e.g.: job change)
- **Non-repetitive** – may not occur more than once in a career (e.g.: death)

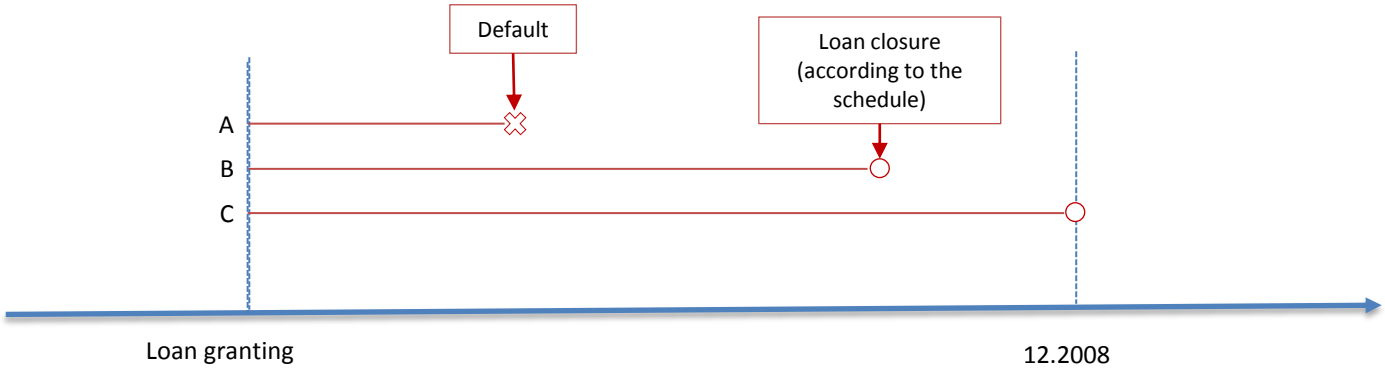
Credit risk:

✓ **Initial event:** granting a loan

✓ Following events are **non-repetitive**: customer that experienced default is assumed not reliable and therefore marked as a bad customer.

Censoring in Credit risk

In the database used we have only right censoring. Therefore we assume equivalent situation when a loan closes in accordance to schedule and a loan that is open at the end of observation period.



Time of origin
Start and end of the observation period must be precisely specified.

Origin of an episode may be:
- Evident/natural – in the study of deaths natural origin of time is birth
- Arbitrary – in the study of retirement age, origin time might be birth or date of entering the labor market.

Credit risk
In credit risk, we assume origin of episode would be date of granting a loan.

Database description

Database description

Database consists of two tables:

1. Production – database about client and transaction (as at the date of application)
2. Transactions – database about history of repayments

Observation unit: trasaction (in other words loan, credit account, contract)

- Assumption: if a client has more than one loan in our bank, event of default on each of his/her loans is independent.

IDs:

- AID – account ID and a key to join two tables
- CID – customer ID

Products:

- CSS: cash loan, for any purpose
- INS: instalment loan, for purchase of particular event

Database description
Transactions data

	cid	aid	product	period	fin_period	status	due_installments	paid_installments	leftn_installments
1	0000000009	css2006012600002	css	200601	200601	A	0	0	24
2	0000000009	css2006012600002	css	200602	200601	A	1	0	24
3	0000000009	css2006012600002	css	200603	200601	A	1	1	23
4	0000000009	css2006012600002	css	200604	200601	A	0	3	21
5	0000000009	css2006012600002	css	200605	200601	A	0	4	20
6	0000000009	css2006012600002	css	200606	200601	A	0	5	19
7	0000000009	css2006012600002	css	200607	200601	A	1	5	19
8	0000000009	css2006012600002	css	200608	200601	A	1	6	18
9	0000000009	css2006012600002	css	200609	200601	A	0	8	16
10	0000000009	css2006012600002	css	200610	200601	A	0	9	15
11	0000000009	css2006012600002	css	200611	200601	A	0	10	14
12	0000000009	css2006012600002	css	200612	200601	A	1	10	14
13	0000000009	css2006012600002	css	200701	200601	A	2	10	14
14	0000000009	css2006012600002	css	200702	200601	A	0	13	11
15	0000000009	css2006012600002	css	200703	200601	A	1	13	11
16	0000000009	css2006012600002	css	200704	200601	A	2	13	11
17	0000000009	css2006012600002	css	200705	200601	A	1	15	9
18	0000000009	css2006012600002	css	200706	200601	A	1	16	8
19	0000000009	css2006012600002	css	200707	200601	A	2	16	8
20	0000000009	css2006012600002	css	200708	200601	A	3	16	8
21	0000000009	css2006012600002	css	200709	200601	A	4	16	8
22	0000000009	css2006012600002	css	200710	200601	A	5	16	8
23	0000000009	css2006012600002	css	200711	200601	A	6	16	8
24	0000000009	css2006012600002	css	200712	200601	A	7	16	8
25	0000000009	css2006012600002	css	200801	200601	B	7	16	8

- Status A/B/C = active / bad / closed
- Due_installments = number of due instalments
- Paid_installments = number of paid instalments
- Leftn_installments = number of instalments left to be paid

Example:

First instalment for transaction ID = css2006012600002 was not paid, subsequent was paid. In April 2006 client paid the instalment that was due as well as the sheduled instalment. In August 2007 defaulted.

	cid	aid	product	period	act_age	act_cc	act_loaninc	app_income
1	0000001237	css2006010100...	css	200601	84	0.5776850886	2.6068821689	1918
2	0000003380	css2006010100...	css	200601	87	0.4997470916	2.5290844714	1977
3	0000003467	css2006010100...	css	200601	75	0.5841584158	4.1254125413	1212
4	0000003620	css2006010100...	css	200601	94	0.6026871401	4.7984644914	1042
5	0000003621	css2006010100...	css	200601	70	0.6045918367	3.1887755102	1568
6	0000003779	css2006010100...	css	200601	97	0.4707766213	4.003202562	1249
7	0000003940	css2006010100...	css	200601	84	0.5311004785	2.990430622	1672
8	0000004025	css2006010100...	css	200601	75	0.5112262522	2.8785261946	1737
9	0000005240	css2006010100...	css	200601	85	0.8916666667	10.416666667	480
10	0000007621	css2006010100...	css	200601	82	0.6264441592	6.4184852375	779
11	0000007991	css2006010100...	css	200601	72	0.6336357292	3.4891835311	1433
12	0000008933	css2006010100...	css	200601	83	0.4875701684	4.0096230954	1247
13	0000009195	css2006010100...	css	200601	79	0.3469761759	3.0543677459	1637
14	0000009992	css2006010100...	css	200601	84	0.5684210526	2.3923444976	2090
15	0000011195	css2006010101...	css	200601	85	0.5816993464	2.7233115468	1836
16	0000014116	css2006010101...	css	200601	86	0.6901098901	5.4945054945	910
17	0000014174	css2006010101...	css	200601	81	0.3948051948	3.2467532468	1540
18	0000014482	css2006010101...	css	200601	93	0.5685950413	4.132231405	1210
19	0000014618	css2006010101...	css	200601	82	0.5931800159	3.9651070579	1261
20	0000014938	css2006010101...	css	200601	76	0.6521181001	6.4184852375	779

1 row = 1 transaction

4 groups of variables:

1. App_: application variables
2. Act_: variables containing current value as at the date of granting a loan
3. Ags_: behavioral variables calculated independently of the customer relations with a bank
4. Agr_: behavioral variables calculated only when a client was a bank's customer for the whole period

```
data trans;
set trans;
by aid cid;
default=(due_installments>=3);
if due_installments>=3 then ttd=intck('month',input(fin_period,yymmn6.),input(period,yymmn6.));
run;

proc sql;
create table trans as
select distinct aid, cid, product, fin_period,
       max(default) as default,
       case when min(leftn_installments)=0 and max(default)=0 then 0
            when min(leftn_installments)>0 and max(default)=0 and max(period)='200812' then 1
            when max(default)=1 then 2 end as rc,
       min(case when ttd >= 0 then ttd end) as ttd,
       case when max(default)=0 then max(period) else min(case when default=1 then period end) end as max_period
  from trans
 group by aid, cid;
quit;
```

Loan closure (according to the schedule)

No default, loan still open at the end of observation period

Default during observation period

SGH

Thank you. Any questions?