Analysis of a commersial register

Per Weidenman
PAR AB

Database

- A collection of data
- It belongs together
- It models the ”world”

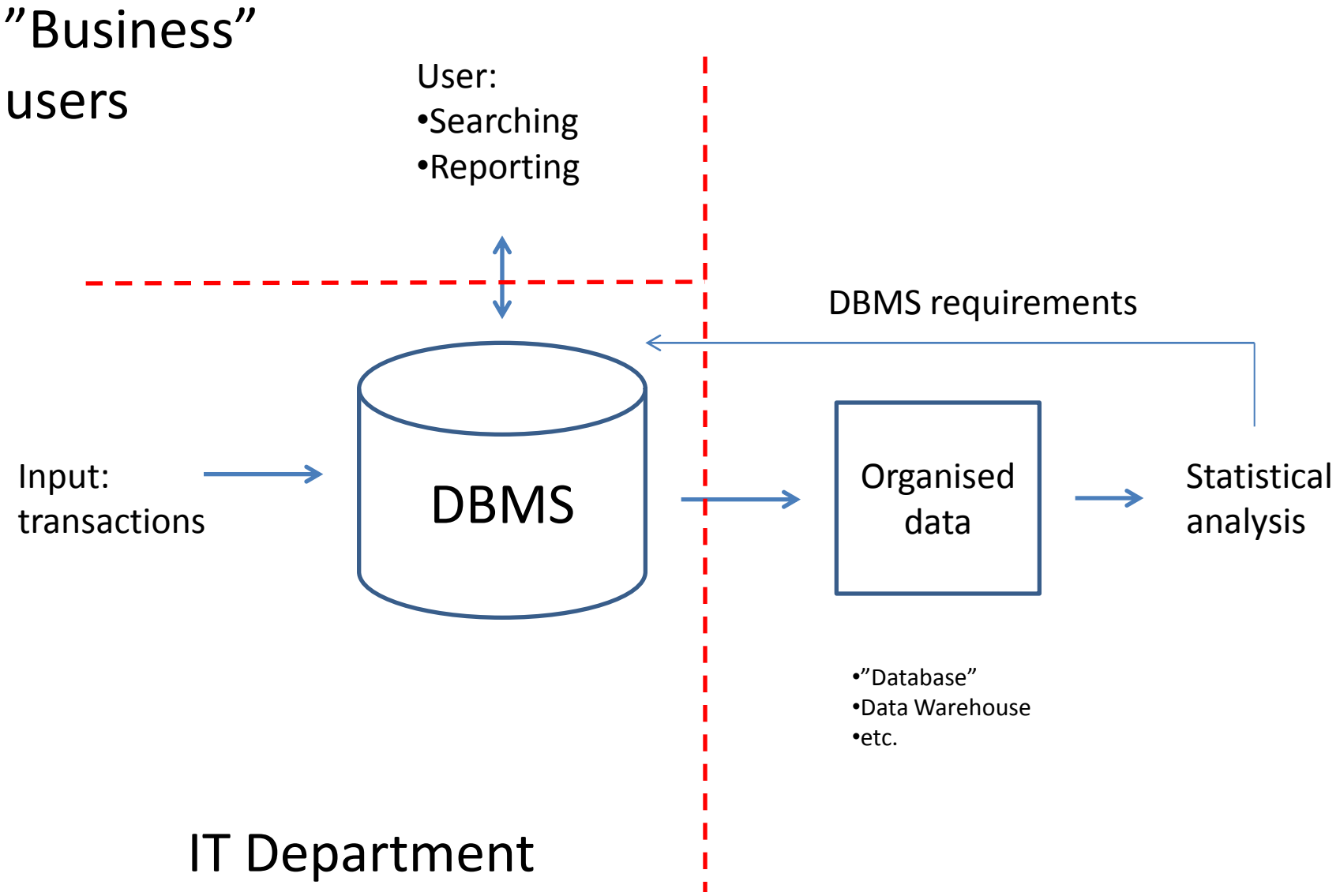
Database management system (DBMS)

- The database (a collection of interrelated data)
- Software to manage and access the data

Relational databases

Tables containing data, organised in rows and columns

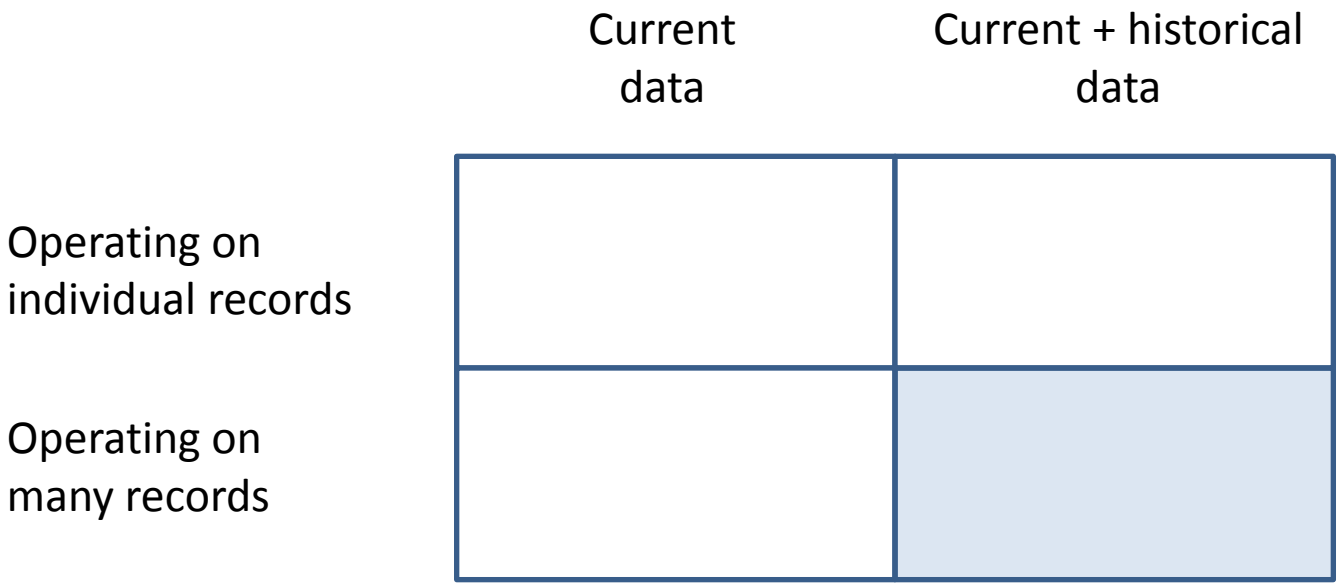
Keys, used for linking data in different tables


”Business”
users
User:
•Searching
•Reporting
DBMS requirements
Input:
transactions
DBMS
Organised
data
Statistical
analysis
•”Database”
•Data Warehouse
•etc.
IT Department

DBMS requirements from a statistical / analytical viewpoint

- Data quality
- Data types
- Performance
- Maximun information
- Historical data
- Regulation and secrecy

DBMS requirements from a statistical / analytical viewpoint



PAR / Bisnode database

Tables

Basic company data
One record per company.
Contains name, address, startdate, enddate, line of business, etc.

FTG

Historic company data
Many records per company.
Contains the accumulated historic records from table FTG

FTG_H

Balance sheet data
One record per annual report (thus many records per company).
Turnover, profit, key ratios, etc.

BOKSLUT

Board member data
Many records per company and person.

FUNKTION_PERIOD

And many more tables!

How?

Serrano

Historic names etc.

Board data

Sampling for times series statistics

Statistical analysis

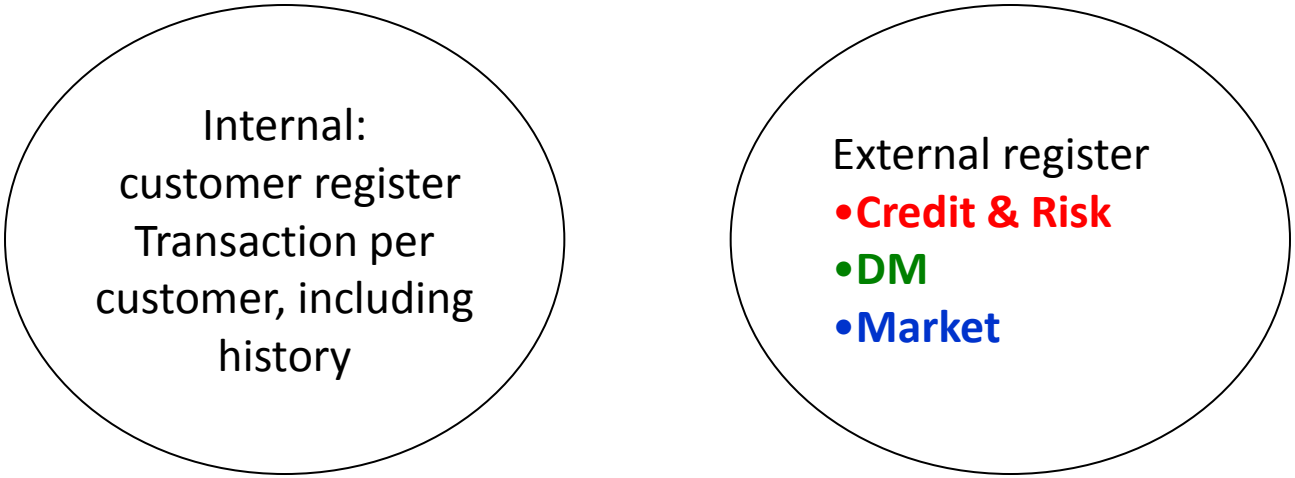
Analysis based on data from commersial registers

From a company point of view

- Internal data
- Mix of internal and external data

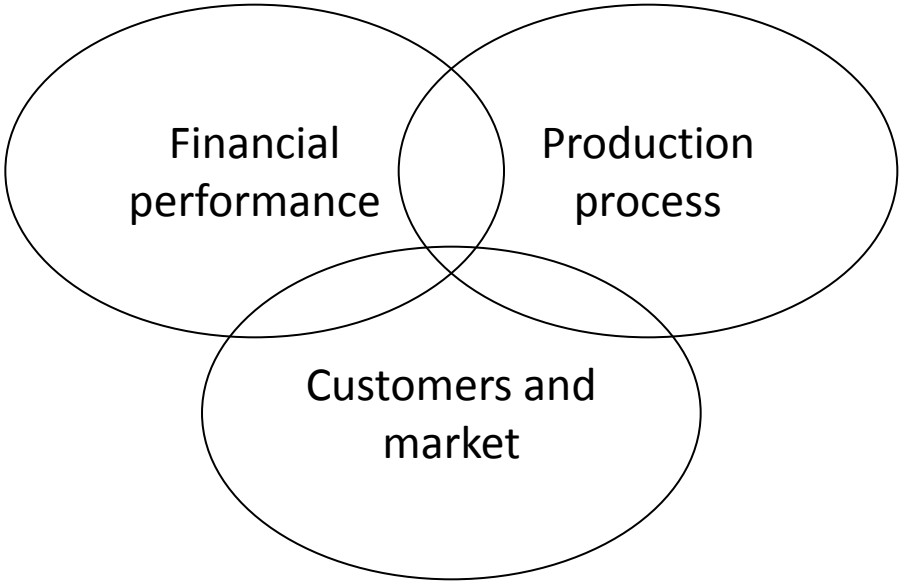
Register use from a company point of view

Useful registers



Register use from a company point of view

Analytics: fields of application for
increased competetive power



Analysis based on a companys internal data

Focus: customers and market

Finding new customers
Planning / executing sales and marketing
Reducing risk – customer credits
Business Intelligence: markets and competitors
.

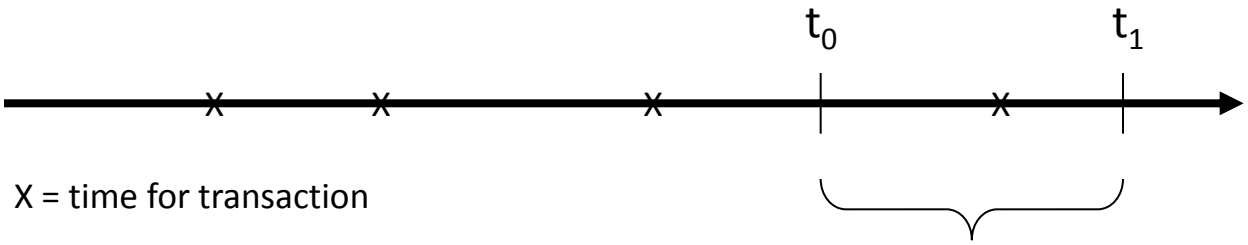
Analysis based on a companys internal data

Focus: customers and market

Objective:
Estimate probability for repeat business > optimize sales in different segments
Estimate Customer Lifetime Value > calculate the financial value of the customer base
Input to company valuation models
.

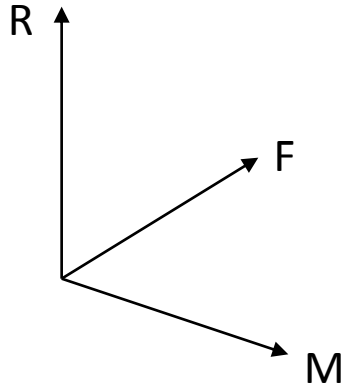
Analysis based on a companys internal data

RFM-models



R = time to last transaction before t_0
F = number of transactions before t_0
M = sum of money from earlier transactions

Elementary use:



Analysis based on a companys internal data

RFM-modeller

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), ”Counting Your Customers: Who They Are and What Will They Do Next?” *Management Science*, 33 (January), 1-24

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2004), ”Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model”, *Market Science* (forthcoming)

Colombo, Richard, and Jiang Weina (1998), ”A Stochastic RFM Model”, *Journal of Interactive Marketing*, Volume 13 (Summer) 2-12

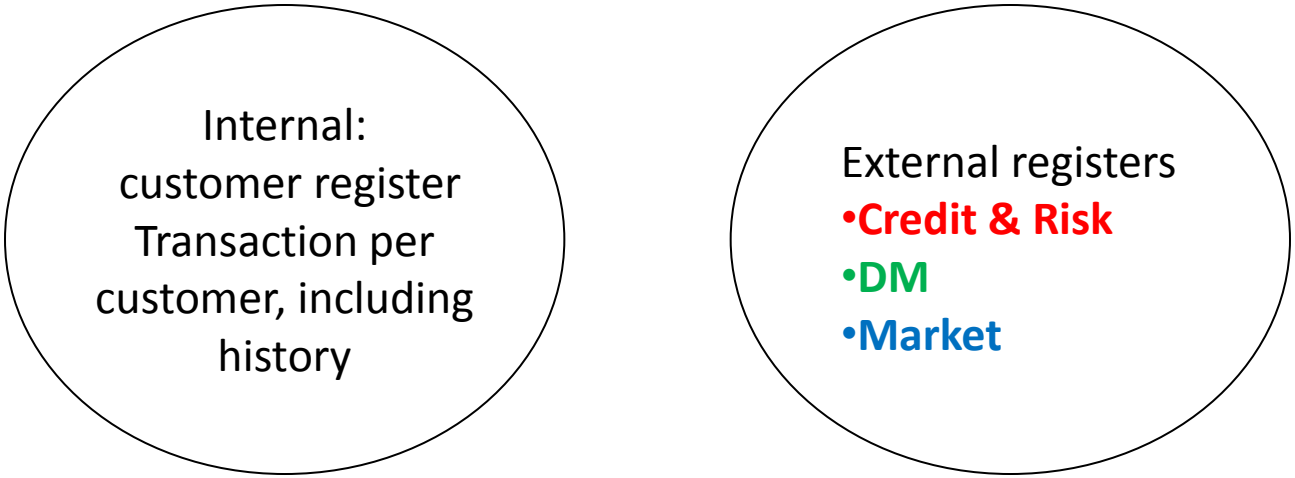
Pfeifer, Phillip E., and Carraway, Robert L., ”Modeling Customer Relationships as Markov Chains”, *Journal of Interactive Marketing*, Volume 14 (Spring 2000) 43-55

Exampel: InkClub

More…

Analysis based on a internal and external data

Useful registers



Analysis based on a internal and external data

External registers

Credit & Risk

DM

Market

Probability for bankruptcy based on stock market data

Probability for bankruptcy based on accounting figures

Altman, E., (1968): ”Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy”, *Journal of Finance*, 23, 689-609

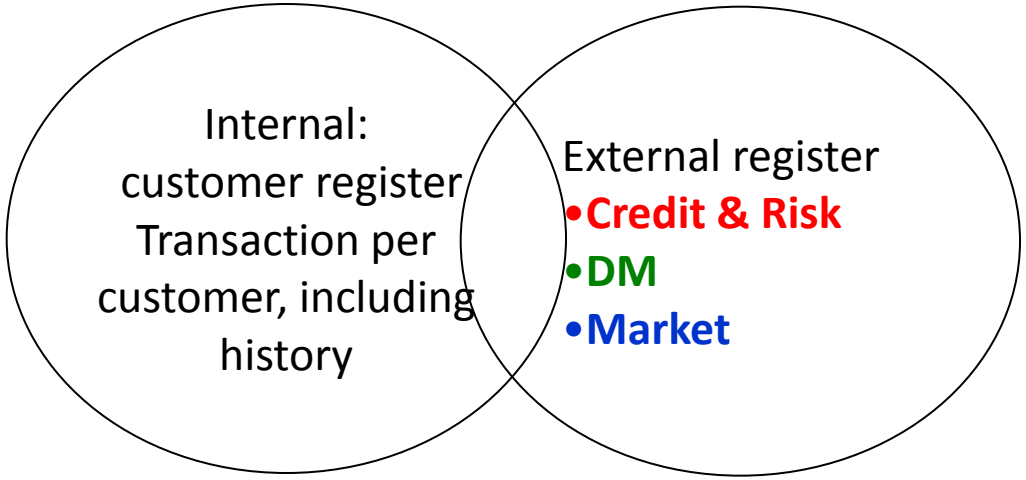
Ohlson, J., (1980): ”Financial Ratios and the Probabilistic Prediction of Bankruptcy”, Journal of Accounting Research, 19, 109-131

’Demografic metod’, modelling lifetime using proportional hazard / Cox regression

Moodys’s riskmodel: scoorecard

Analysis based on a internal and external data

Combination of internal and external registers

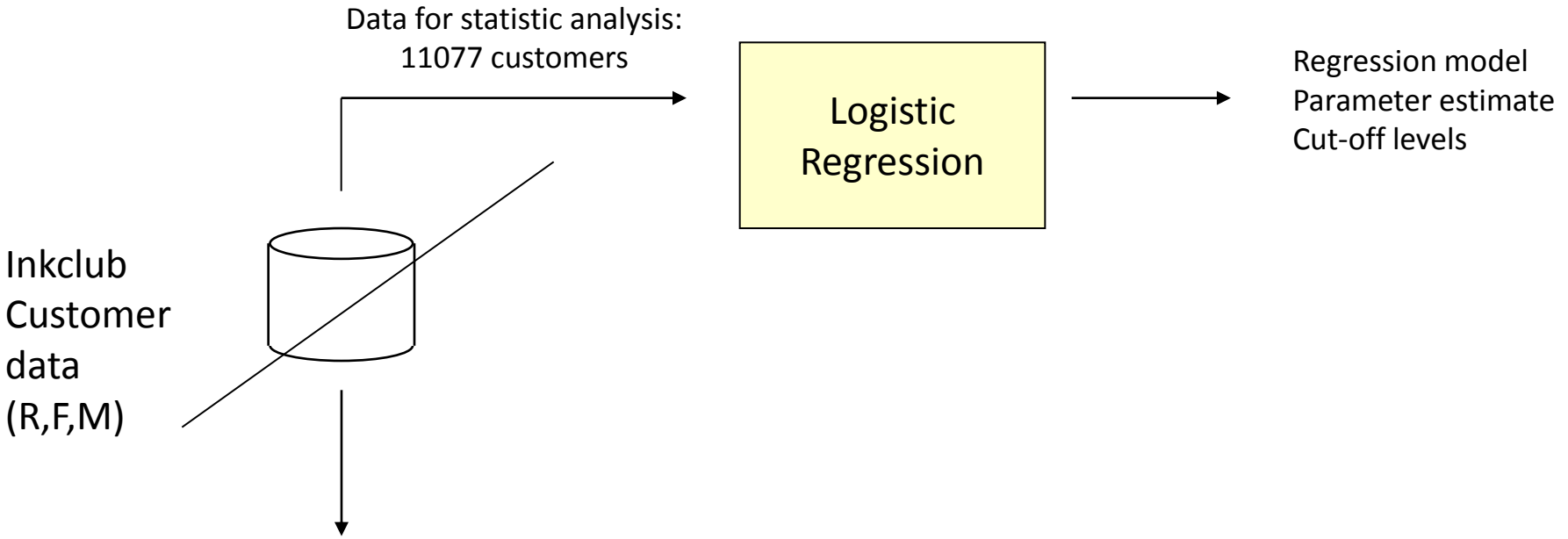


Analysis based on a internal and external data

Different approaches
- ”Automated” modeling
- Neural networks
- Traditional statistics

Prediction model
Firm of accountants

END


Data for statistic analysis:
11077 customers
Logistic
Regression
Regression model
Parameter estimate
Cut-off levels
Inkclub
Customer
data
(R,F,M)

Data for statistic analysis:
11077 customers

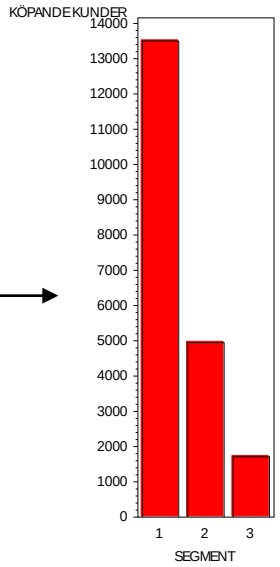
Logistic Regression

Regression model
Parameter estimate
Cut-off levels

Inkclub Customer data (R,F,M)

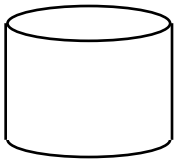
Data for evaluation
11246 customers during12 months:
~ 118000 observations

Chosen Regression model



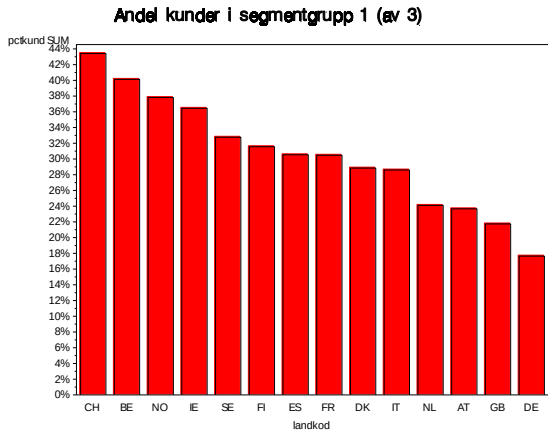
$$\text{logit}(p) = \beta_0 + \beta_1 \ln(R) + \beta_2 \ln(F) + \beta_3 \ln(M)$$

Inkclub Customer data (R,F,M)

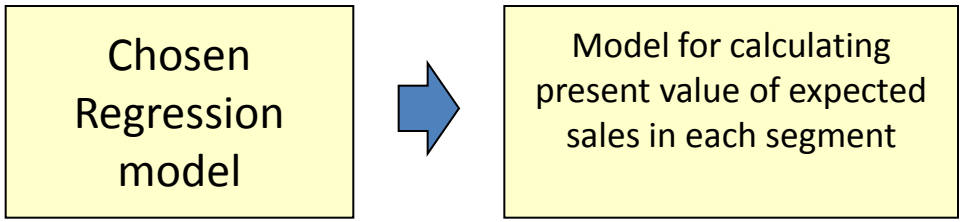


Data for evaluation
11246 customers during12 months:
~ 118000 observations

Chosen Regression model



$$\text{logit}(p) = \beta_0 + \beta_1 \ln(R) + \beta_2 \ln(F) + \beta_3 \ln(M)$$



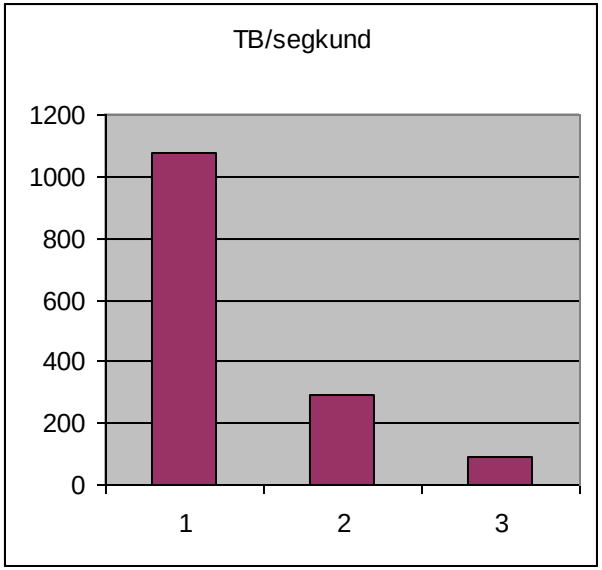
Segmenting all customers based on RFM-data

Calculated present value based on estimated

- Customer lifetime
- Order frequency
- Income per order

Chosen Regression model

Model for calculating present value of expected sales in each segment



Segmenting all customers based on RFM-data

Calculated present value based on estimated
- Customer lifetime
- Order frequency
- Income per order

Expected present value per customer in three RFM-segment

RFM

Recency, Frequency, Money

Purpose:

Predicting future buying pattern and customer lifetime by analysing earlier buying behaviour

Input to CLV models for better modeling of Customer Lifetime Value

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), ”Counting Your Customers: Who They Are and What Will They Do Next?” *Management Science*, 33 (January), 1-24

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2004), ”Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model”, *Market Science* (forthcoming)

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2004), ”RFM and CLV: Using Iso-Value Curves for Customer Base Analysis” *Journal of Marketing Research*.

Reinartz, Werner J., V. Kumar (2000), ”Customer Lifetime Duration: An Empirical Framework For Measurement and Explanation”, INSEAD Working Paper, August 2000

CLV

Customer Lifetime Value

Purpose:

For every individual customer: calculating the discounted value of all future profits measured over the customers life cycle

$$CLV_i = \sum_{t=0}^{T} f(\text{revenue}_{it}, \text{cost}_{it}, \text{retention rate}_{it}, \text{discount factor}_t)$$

Bauer, Hans H., Maik Hammerschmidt, and Matthias Braehler (2003), ”The Customer Lifetime Value Concept and its Contribution to Corporate Valuation, *Yearbook of Marketing and Consumer Research*, Vol. 1

Rosset, S., E. Neumann, U. Eick, N. Vatnik (2002), ”Customer Lifetime Value Models for Decision Support”, *Data Mining and Knowledge Discovery*, 7, 321-339

Werner, Uta (2003) ”Getting Real About Customer Lifetime Value”, Marakon Associates

Customer Equity

Purpose:

Transforming the operative CLV measure to an instrument for valuation of the company

Instrument for changing view: from product focus to market focus [from ’Brand equity’ to ’Customer equity’]

Blattberger, R. C., Deighton J. (1996), ”Manage Marketing by Customer Equity”, *Harvard Business Review*, Vol. 74

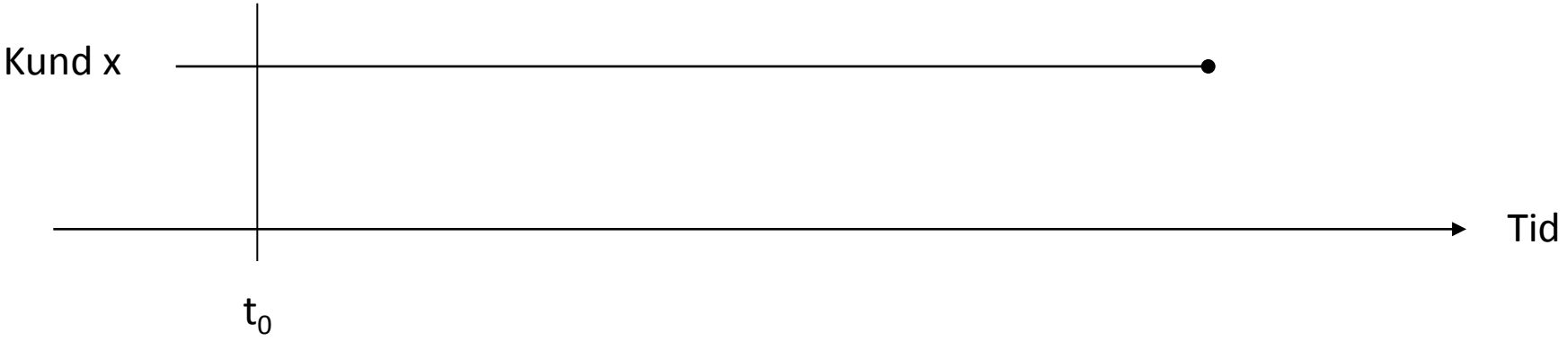
Bauer, Hans H., Maik Hammerschmidt, and Matthias Braehler (2003), ”The Customer Lifetime Value Concept and its Contribution to Corporate Valuation, *Yearbook of Marketing and Consumer Research*, Vol. 1

Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, Gordon Swartz (2002), ”Seven Barriers to Customer Equity Management”, *Journal of Service Research*, Vol. 5 (August)

Pareto/NBD model for customer behaviour

The number of transactions during time interval t is Poisson with transaction rate λ

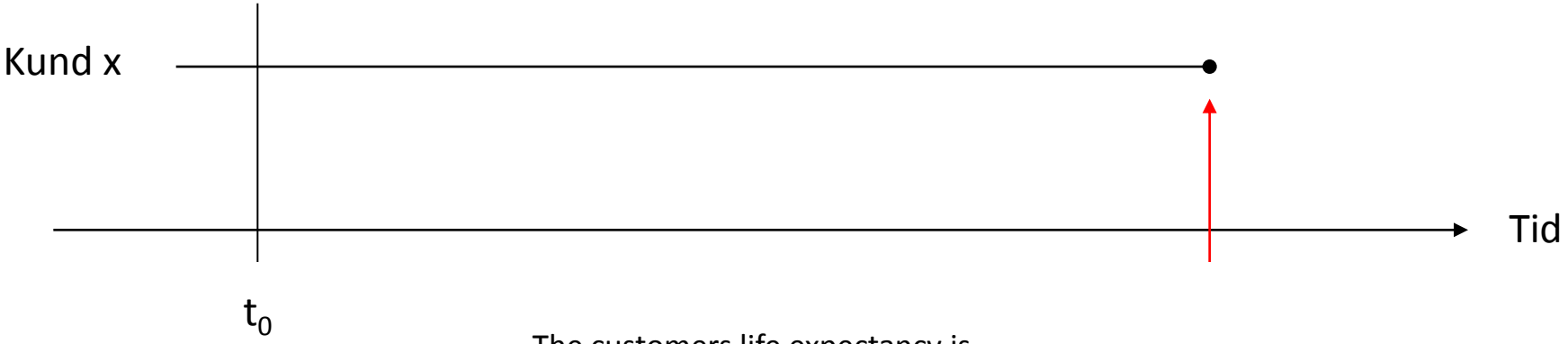
Over all customers λ is gamma distributed with parameters r and α



Pareto/NBD model for customer behaviour

The number of transactions during time interval t is Poisson with transaction rate λ

Over all customers λ is gamma distributed with parameters r and α

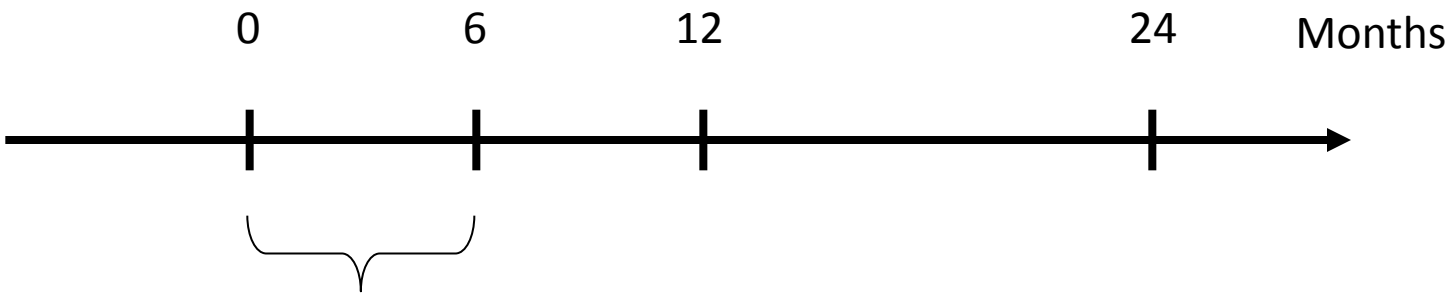


The customers life expectancy is Exponentially distributed with dropout rate μ

Over all customers μ is gamma distributed with parameters s and β

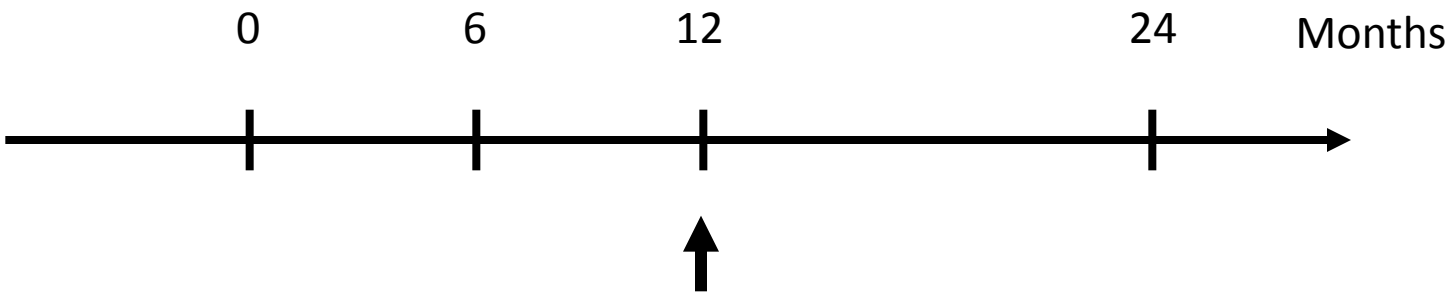
λ and μ are independent

Prediction model for buying behaviour



A sales activity has generated a number of new customers during the first six months.

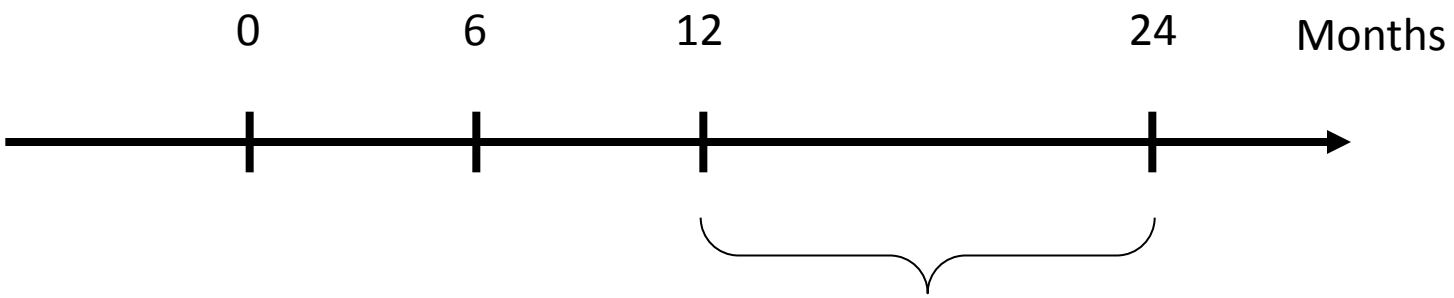
Prediction model for buying behaviour



At the end of the year we need to predict: how will repeat business look like during the coming year?

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2004), ”Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model”, *Market Science*.

Prediction model for buying behaviour



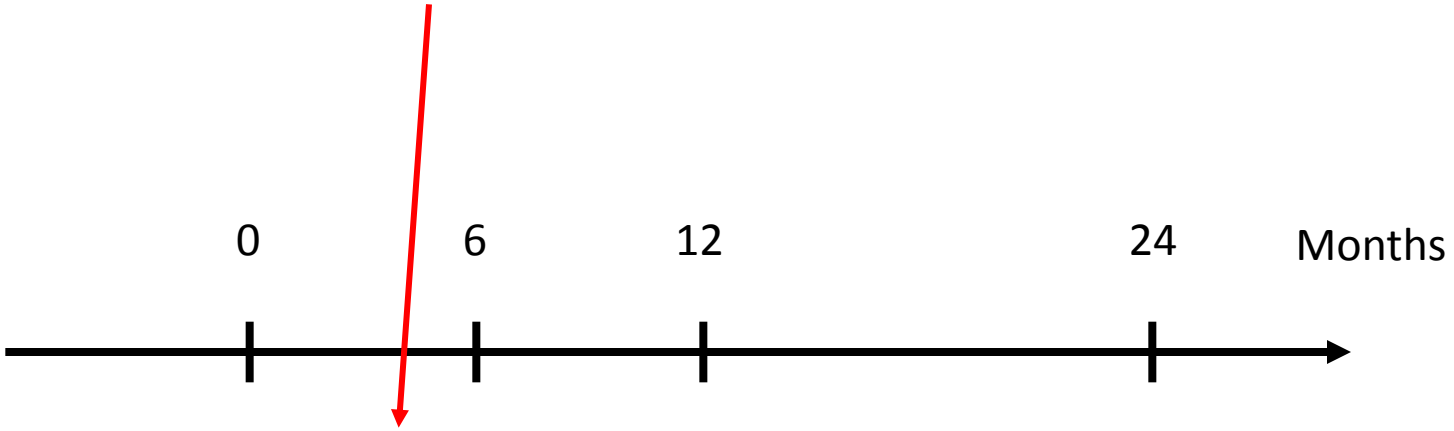
The prediction will give us expected buying frequency during chosen period.

Multiply with expected sales
sum > CLV for new customers.

Prediction model for buying behaviour

Input data for the model:

- Number of additional sales during first 12 months.



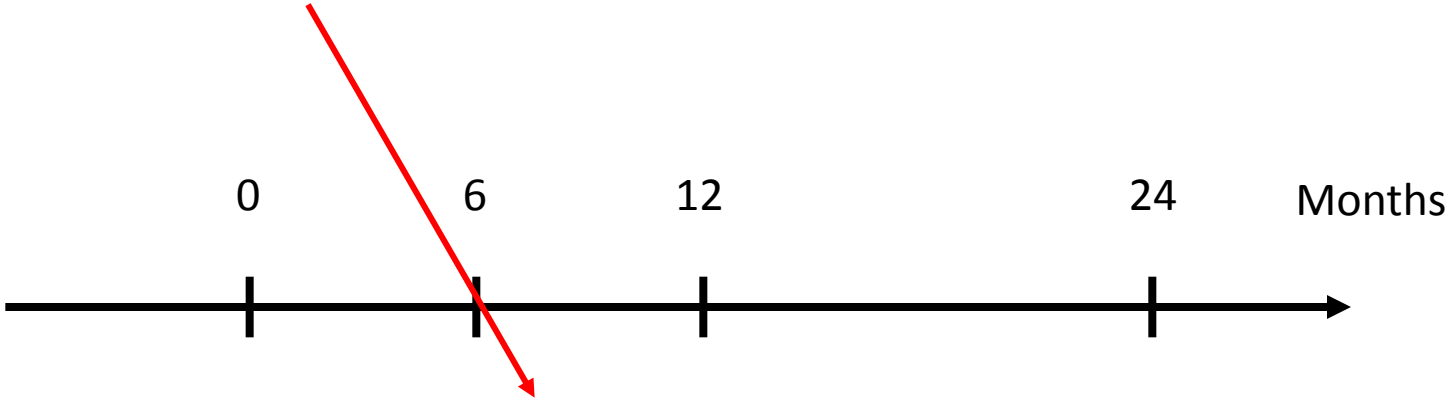
Data concerning 1500 new air freight customers during2001.

ID	x	t_x	T
734	1	29	49,57143
735	0	0	27,57143
736	0	0	28,42857
737	0	0	48,57143
738	0	0	51,85714
739	1	0,142857	49
740	0	0	33,42857
741	0	0	29,57143
742	1	0,857143	35,57143

Prediction model for buying behaviour

Input data for the model:

- Number of additional sales during first 12 months.
- Time for last transaction

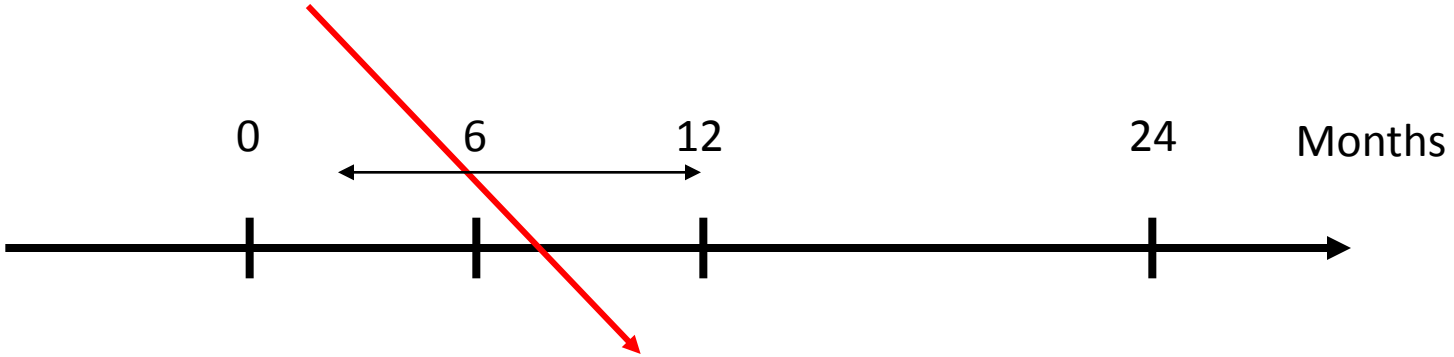


ID	x	t_x	T
734	1	29	49,57143
735	0	0	27,57143
736	0	0	28,42857
737	0	0	48,57143
738	0	0	51,85714
739	1	0,142857	49
740	0	0	33,42857
741	0	0	29,57143
742	1	0,857143	35,57143

Prediction model for buying behaviour

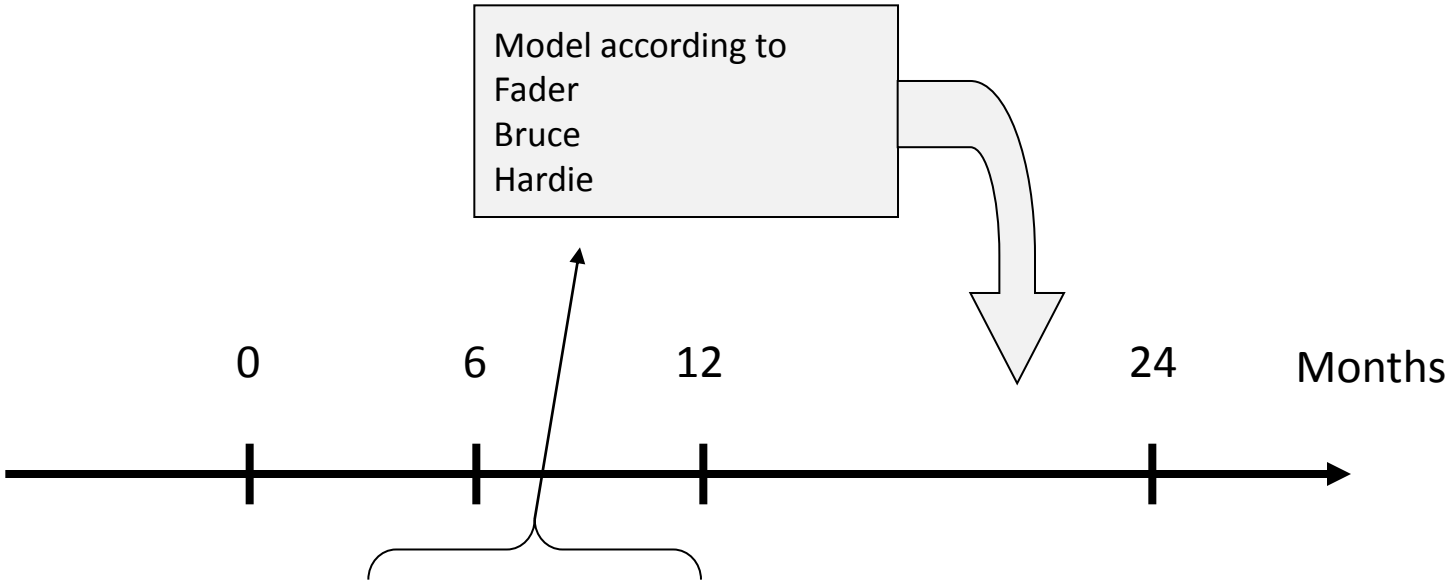
Input data for the model:

- Number of additional sales during first 12 months.
- Time for last transaction
- Length of customer relationship



ID	x	t_x	T
734	1	29	49,57143
735	0	0	27,57143
736	0	0	28,42857
737	0	0	48,57143
738	0	0	51,85714
739	1	0,142857	49
740	0	0	33,42857
741	0	0	29,57143
742	1	0,857143	35,57143

Prediction model for buying behaviour

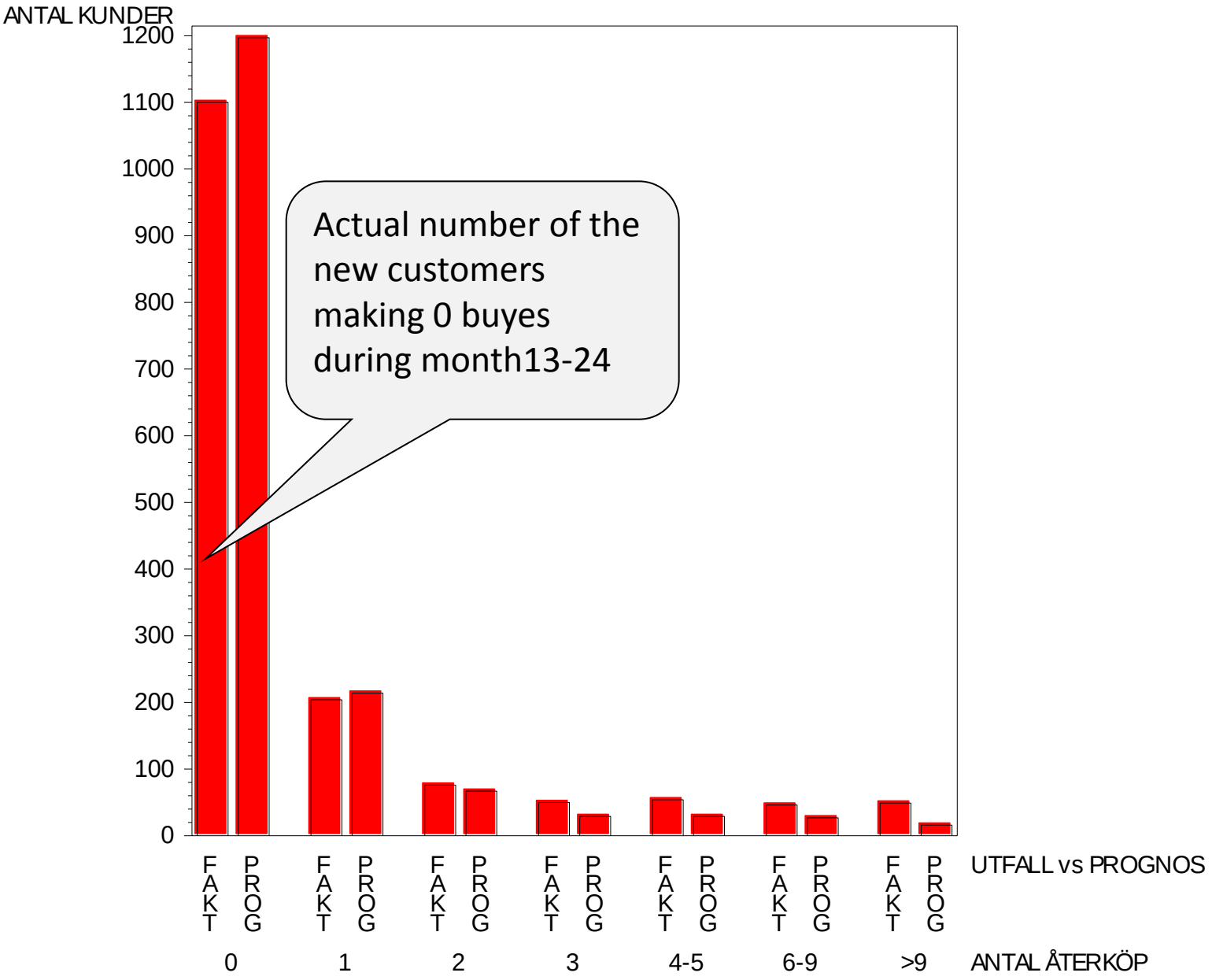


ID	x	t_x	T
734	1	29	49,57143
735	0	0	27,57143
736	0	0	28,42857
737	0	0	48,57143
738	0	0	51,85714
739	1	0,142857	49
740	0	0	33,42857
741	0	0	29,57143
742	1	0,857143	35,57143

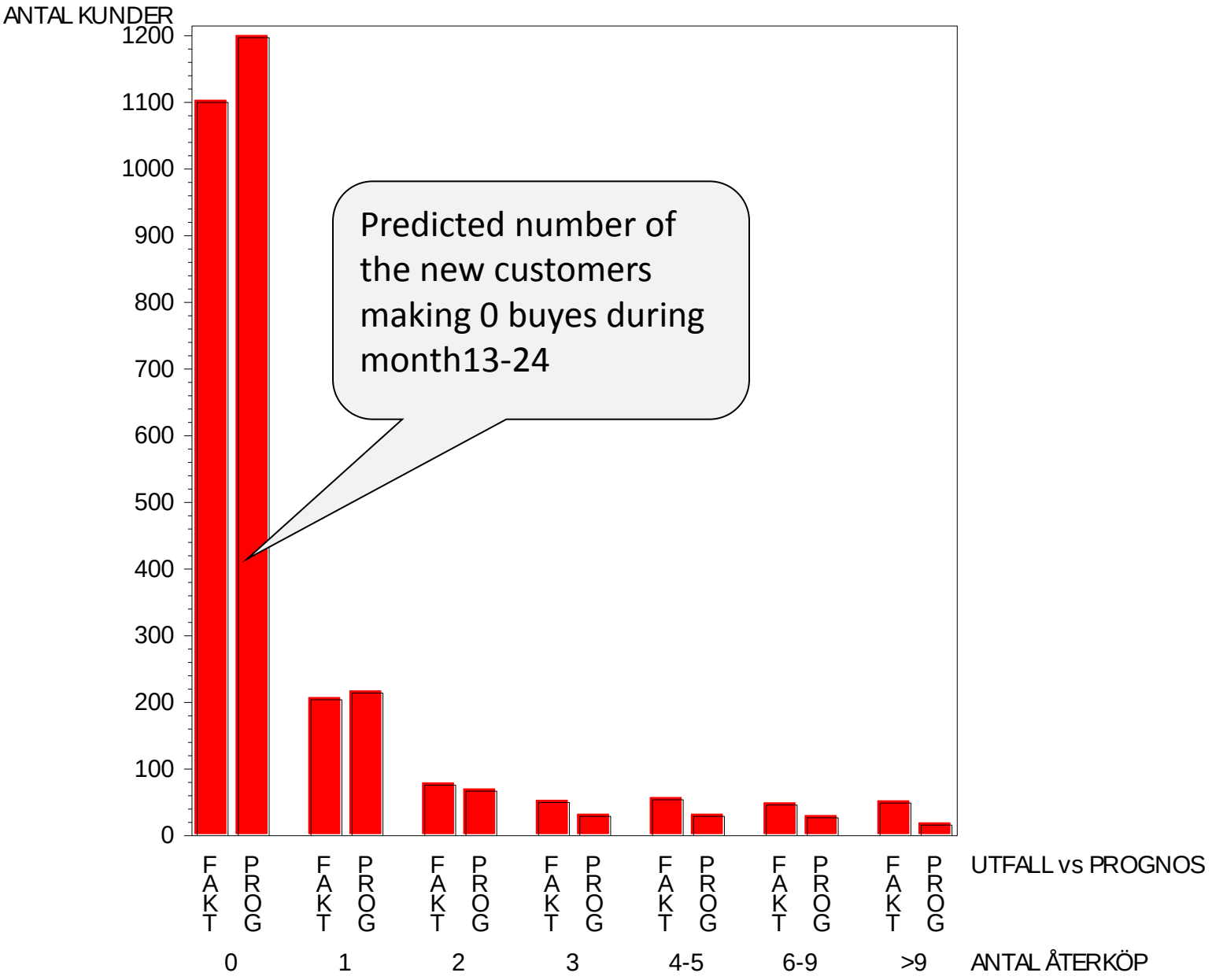
Prediction, compare

- Actual buying frequency
- Predicted buying frequency

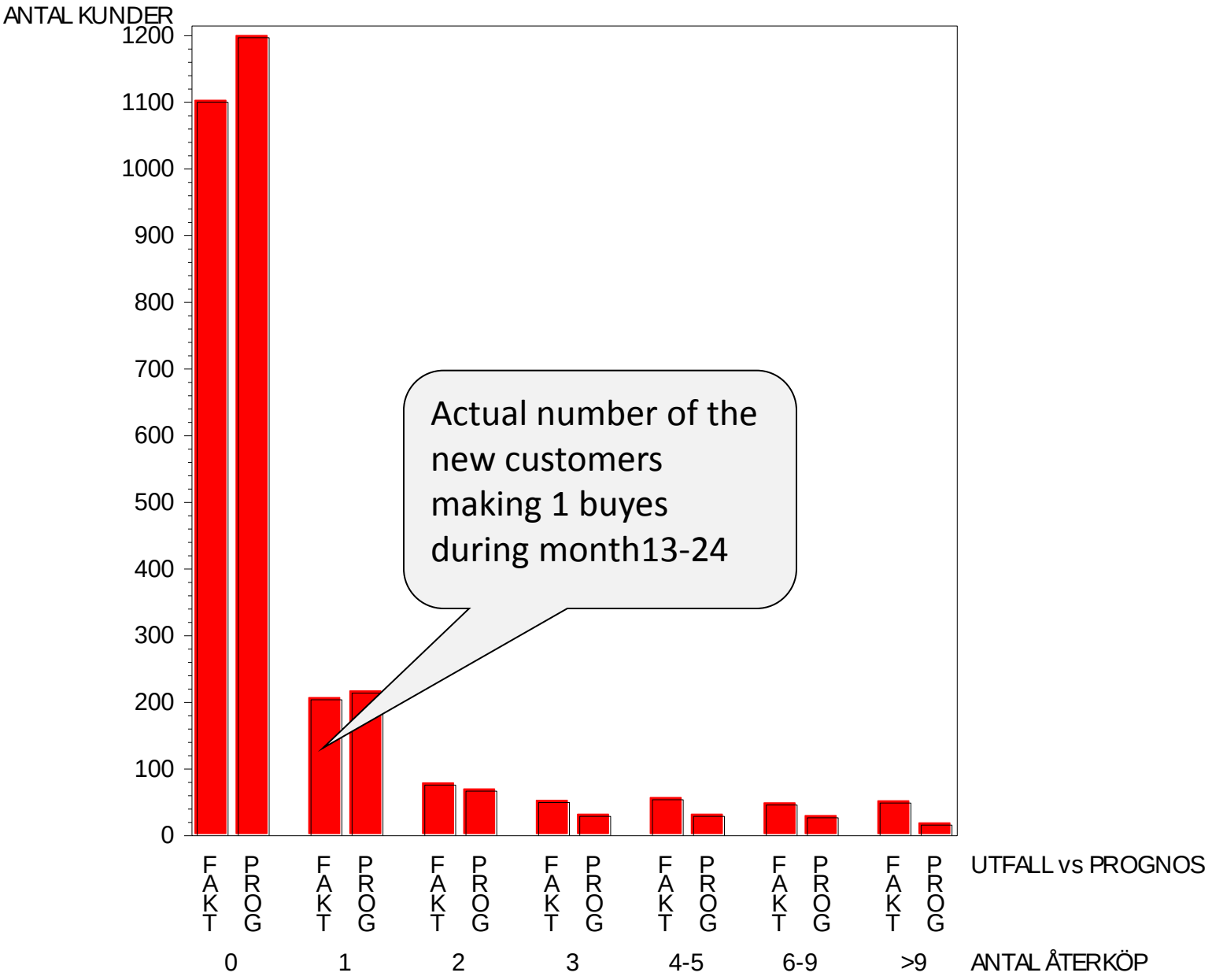
Prediction model for buying behaviour



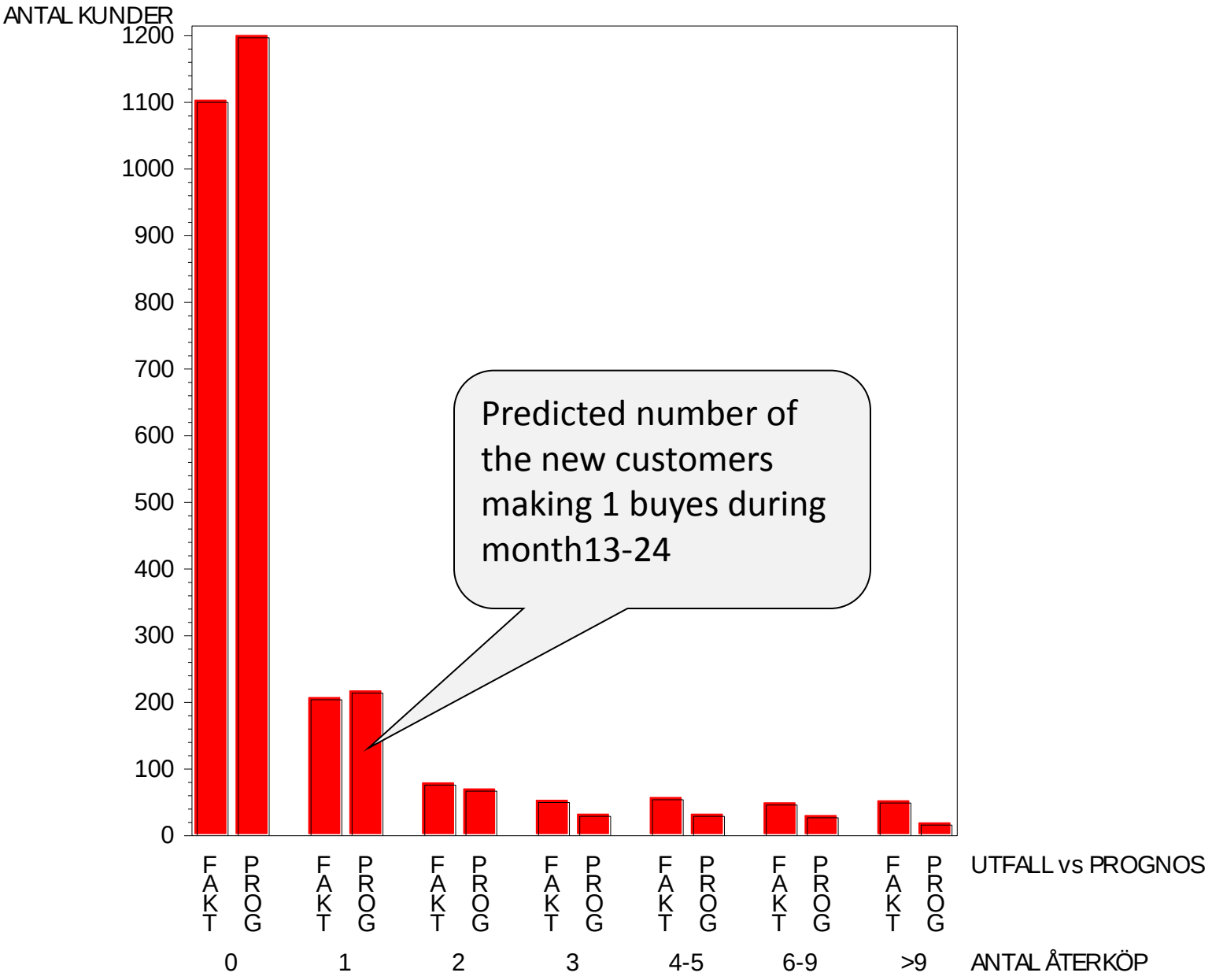
Prediction model for buying behaviour



Prediction model for buying behaviour



Prediction model for buying behaviour



Prediction model for buying behaviour

Read more about predicting buying behaviour based on transaction data:

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2004), ”Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model”, *Market Science*.

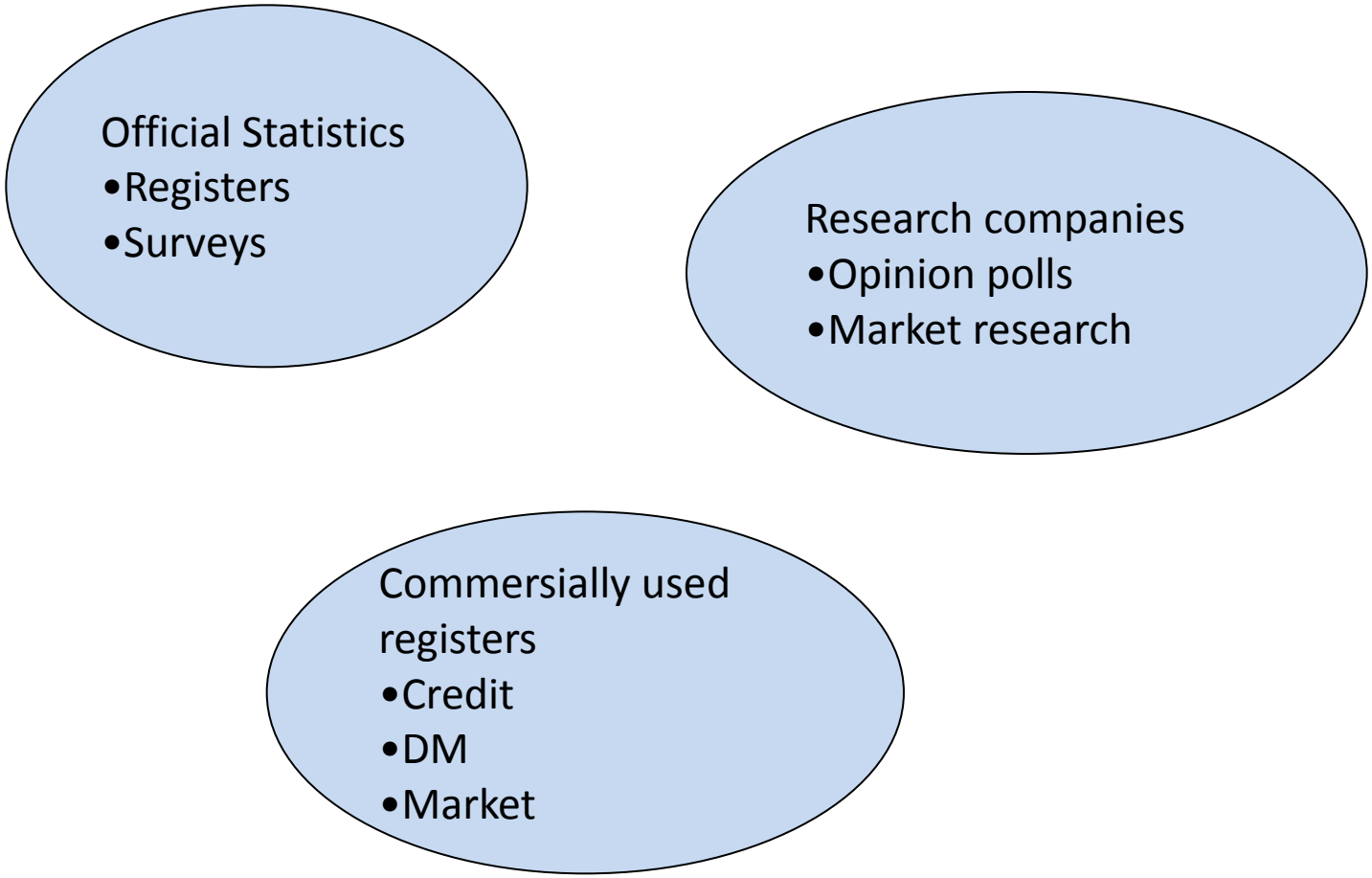
Fader, Peter S., Hardie (2001), ”Forecasting Repeat Sales at CDNOW: A Case Study”, *Interfaces*.

Fader, Peter S., Hardie, and Ka Lok Lee (2004), ”RFM and CLV: Using Iso-Value Curves for Customer Base Analysis”, *Journal of Marketing Research*.

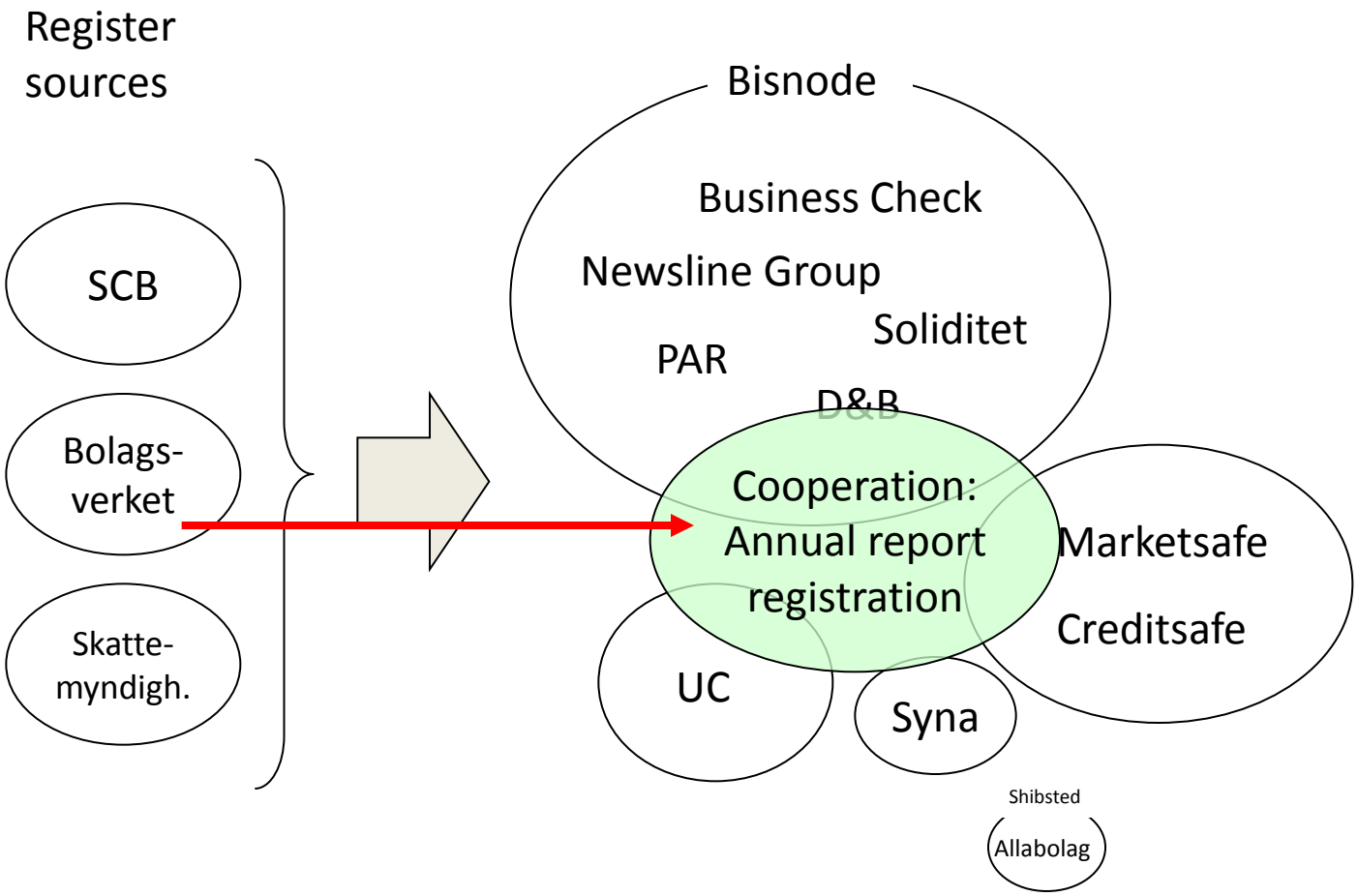
Fader, Peter S., Hardie, and Ka Lok Lee (2004), ”Implementing the BG/NBD Model for Customer Base Analysis in Excel”.

Excel sheet containing the calculations:

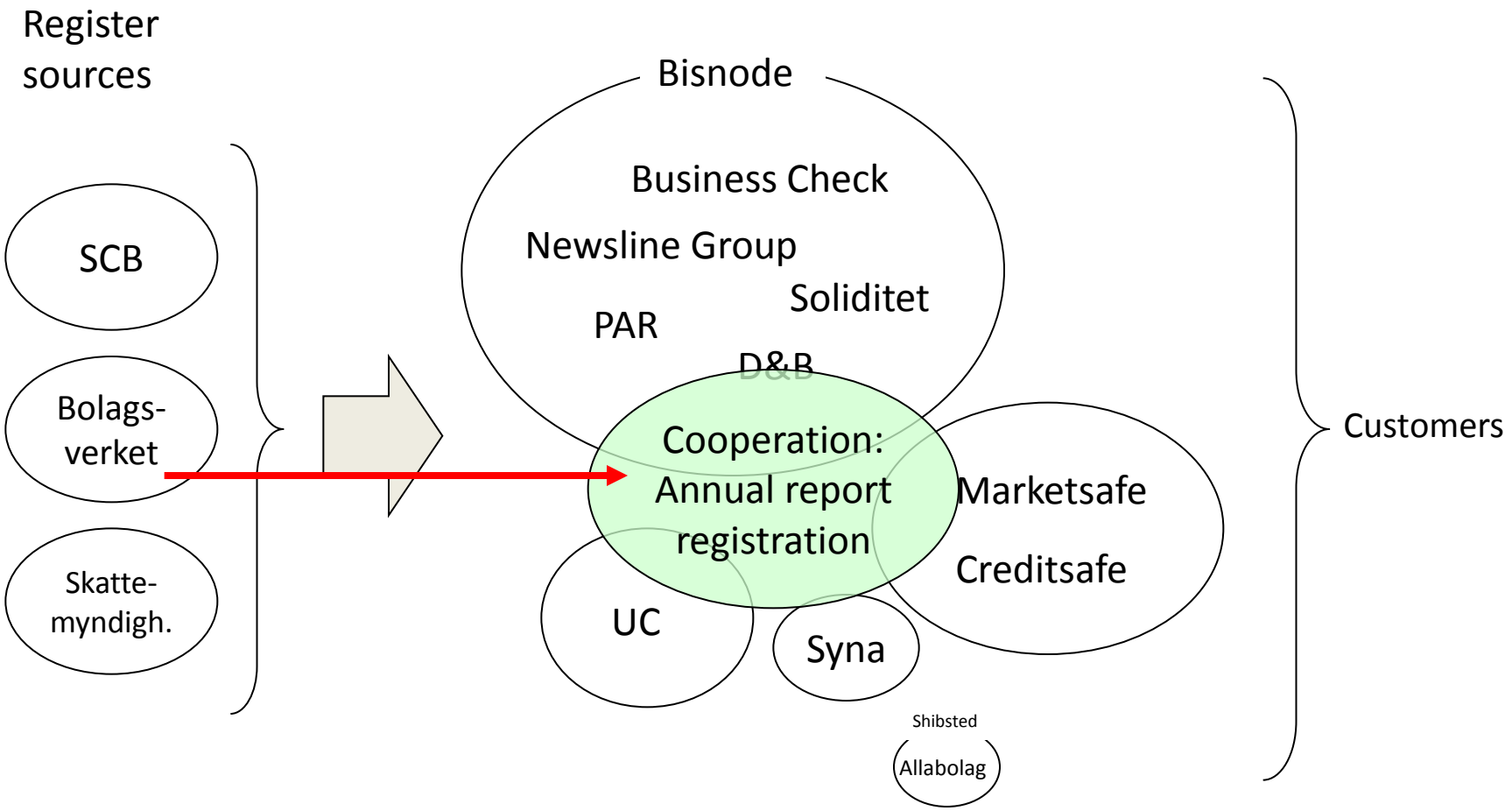
http://www.brucehardie.com/notes/004/


Official Statistics
•Registers
•Surveys
Research companies
•Opinion polls
•Market research
Commersially used registers
•Credit
•DM
•Market

Suppliers


Register sources
SCB
Bolags-verket
Skatte-myndigh.
Bisnode
Business Check
Newsline Group
Soliditet
PAR
D&B
Cooperation:
Annual report registration
Marketsafe
Creditsafe
UC
Syna
Shibsted
Allabolag

Suppliers


Register sources
SCB
Bolags-verket
Skatte-myndigh.
Bisnode
Business Check
Newsline Group
Soliditet
PAR
D&B
Cooperation:
Annual report registration
UC
Syna
Marketsafe
Creditsafe
Shibsted
Allabolag
Customers