# Evaluating the effect of model quality in optimisation

Dr. Vladimir Fishman and Dr. Anatoly Reynberg

Credit Scoring and Credit Control XI Conference August 2009

Experian
Decision Analytics

# Topics of presentation

- Customer management optimisation problem
- Mathematical formulation
- Evolution of analytical approaches
- Segment based approach versus record level approach: comparison
- Importance of model quality for record level optimisation
- Multi-goal optimisation
- Role of uncertainty (two types of uncertainty)
- Model quality and its relation to expected value of perfect information
- Evaluating effect of model quality
    - Objectives
    - Functionality
    - Description of optimisation problem
    - Cumulative distributions for different values of model quality measure
    - Role of constraints
    - Taking correlation into account
    - Results for response and profitability models
    - Results for different data segments: comparison of two optimisation tasks
- Conclusion

# Customer management optimisation problem

*Customer Data*

- Response data
- Financial data
- Personal data,

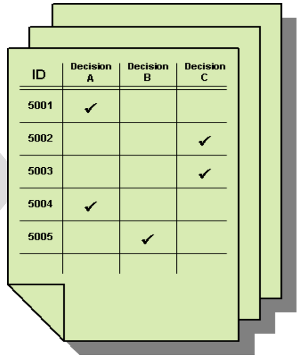
Predictive

Modeling

*Metrics*

Interest revenue
Market penetration
Bad debt losses
Prepayment losses
Application processing
cost

?




*Financial Institution*

*Credit Card Portfolio - 10 M customers*

*Optimal Line assignment problem*

- Selling costs
- Solicitation costs
- Channel coefficients
- Revenues
- Eligibility conditions
- Mandatory triggers

*Decision Options*

- Maximize profit
- Budgetary constraints
- Contact rules

*Business Goals*
*Business Constraints*
*Contact Protocols*

*Decision options:$500, $1000,…$5000*

*Goal: Maximize profit*

*Constraints: Maximum bad debt*

Experian™

# Customer management optimisation problem

# Mathematical formulation

The objective is to maximize utility function:

$$U(X) = \sum_{i=1}^{N}\sum_{j=1}^{M} u_{ij}x_{ij}$$

$$x_{ij} = 0,1$$

Subject to constraints:

Global (resource)

$$\sum_{i=1}^{N}\sum_{j\in J_l}^{M} m_{ij}^{l}x_{ij} < b_l$$

$$l = 1 \ldots L$$

Customer level conditions

$$\sum_{j\in J_k}^{M} s_{ij}^{k}x_{ij} < h_i^k$$

$$k = 1 \ldots K$$

# Customer management optimisation problem: objectives

## Analytical requirements

- Feasibility
- Optimality

**Objective function is maximized (minimized) and**

**all constraints are satisfied**

## Operational requirements

- **Transparency:**

  **Not a "black box", allows for interpretation**

- **Flexibility:**

  **Easy to change objective functions, constraints etc.**

- **Ease of implementation**

  **Fit into commonly used operational processes**

# Customer management optimisation problem: evolution of analytical approaches

- **Empirical**

  Traditional champion/challenger trial and error method. Slow and heavily depends on analysts experience and intuition.

- **Segment based optimisation**

  Aggregate similar customer accounts into clusters using non-supervised learning technique, then treat all accounts in a cluster as identical. Optimisation problem becomes low dimensional and can be solved using standard technique.

- **Record level scalable optimisation**

  Analytically advanced. Delivers precise scalable solution.

- **Hybrid approach**

  Combines record level optimisation with supervised learning technique. More transparent and easier to integrate.

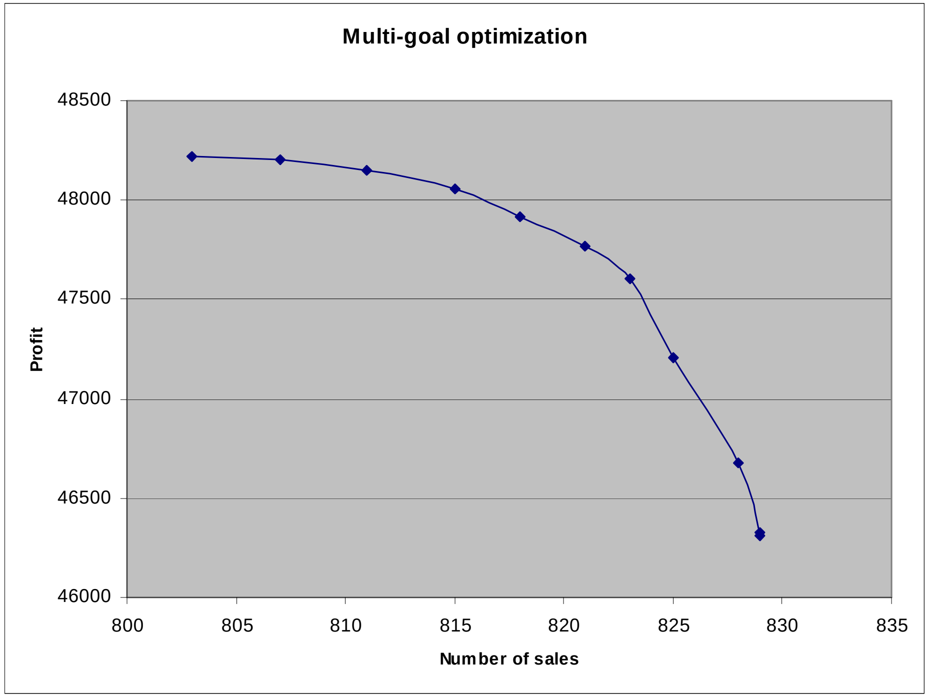
## Segment based approach versus record level approach: comparison
## Importance of model quality for record level optimisation

- If models predicting customer behavior are more sophisticated, predictions will be different for different customers; as a result, data is no longer segmented. In such cases the segmnentation solution becomes significantly sub-optimal.
- Within segment based approach the level of sub-optimality cannot be evaluated. Based on comparison with record level optimisation for real business cases it can be significant (~20%).
- Segment level optimisation method is not flexible: different tasks need different segmentations.
- Since segment-based optimisation treats all customer in a segment as identical, it cannot properly handle even simple record level contact rules and conditions.
- Only record level optimisation can utilize to full extend model predictive power

# Multi-goal optimisation

Most practical optimisation problems in marketing risk management are multi-objective in nature.

To analyze trade-off between conflicting objectives we generate a Pareto set or efficient frontier. The multi-objective curve contains a number of points, representing different compromises between two objectives.



# Taking uncertainty into account: general considerations

- Making optimal decision under uncertainty is a mathematical challenge.
- The issue was never addressed for problems of the size we typically work with (millions of records, hundreds of decision options).
- It is often a multi-stage problem: first initial stage decision should be made, then on the next stage when some uncertainty is resolved one make new decision and so on.
- Let us denote solution (utility function) of stochastic recourse problem RP. If uncertain values are replaced by their expected values, problem becomes deterministic with solution EV. That is what we do in our standard approach when we use expected profitability or response rate without accounting for uncertainty.
- Accounting for uncertainty in general improves utility. This improvement is called *value of stochastic solution* (VSS):
- VSS = RP – EV
- If the uncertainty could be resolved in advanced (if we knew the future outcomes of uncertain variables), the optimal solution would be (it is often called WS: *wait-and-see solution*) in general more profitable and this increase in utility function is called *expected value of perfect information*:
- EVPI = WS - RP

# Taking uncertainty into account: two types of uncertainty

- First type is record level uncertainty regarding individual customer characteristics (profitability, response rate, etc.). If this uncertainty is independent of other customers, we may call it diversifiable uncertainty. Accounting for this type of uncertainty is important in following cases:

  1. Evaluating the effect of model quality

  2. When decisions are triggered by uncertain events on individual level (e.g. collection optimisation)

- Second type is uncertainty in underlying parameters: to develop a successful strategy, we must assess its sensitivity to a number of key characteristics which are known only with some level of uncertainty, and which may change in the future as the economic and demographic environment evolves. Such key characteristics can be interest or inflation rates, average income, etc. In multistage campaigns, where decisions are made on a stage-by-stage basis, at the end of each stage some of these uncertainties are resolved, which means the problem we have to deal with is a high dimensional multiple recourse optimisation problem.

# Taking uncertainty into account: formulation

Utility:

$$\max E_\xi(\sum_{i=1}^{N}\sum_{j=1}^{J} a_{ij}^{\xi} u_{ij}^{\xi}(\xi)) = \max(\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_{\max}(\xi)} a_{ij}^{k} p_{jk} u_{ij}^{k}(\xi))$$

Constraints:

On average value

$$E_\xi(\sum_{i=1}^{N}\sum_{j\in J_m}^{J} a_{ij}^{\xi} v_{ij}^{\xi m}(\xi)) = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_{\max}(\xi)} p_{jk} a_{ij} v_{ij}^{km}(\xi) < C_m^1$$

Probabilistic

$$\Pr(\sum_{i=1}^{N}\sum_{j\in J_m}^{J}\sum_{k=1}^{K_{\max}(\xi)} p_{jk} v_{ji}^{km}(\xi) > C_m^2) < Eps^m$$

Worst case

$$\sum_{i=1}^{N}\sum_{j\in J_m}^{J}\sum_{k=1}^{K_{\max}(\xi)} v_i^m(\mathbf{a_i},\xi) < C_m^3$$

Record level conditions: : arbitrary Boolean expression of ($m$ = 1…M) linear conditions:

$$A_m = \sum_{i\in I_m}^{N_m} \alpha_i a_i - G_m > 0$$

# Impact of model quality analysis: objectives

- Discrepancy between prediction and actual values is a record level diversifiable uncertainty
- In case of linear optimisation problem results cannot be improved by just taking the uncertainty in account without improving model quality (*value of stochastic solution=0*)
- Improving model quality can be costly. Therefore, it is important to assess in advance the impact of increasing model precision on the overall result (profitability, market penetration, etc.) of the chosen strategy. This issue is related to determining EVPI (*expected value of perfect information*). This impact can be significant and depends on choice of utility function and constraints.
- The idea of the method is to add a variation to prediction values in order to simulate a model with a desirable quality, then solve corresponding optimisation problem and evaluate impact of the model quality improvement
- The method helps to create a roadmap for building, improving, and updating predictive models.
- Method provides cost benefit analysis for model improvement efforts

# Impact of model quality analysis: functionality overview

- **Input:**
  - Original individual level prediction values and current optimisation results
  - Actual or simulated output or current measures of model quality
  - Correlation between prediction values
  - Desirable improvement of model quality
- **Output**
  - Expected improvement of optimisation results based on simulated prediction values consistent with desirable model quality and correlation
- **Model quality measures:**

| | |
|---|---|
| - Gini index | binary dependent variable |
| - Kolmogorov-Smirnov criterion value | binary and continuous dependent variable |
| - ROC – receiver operating characteristic | binary dependent variable |
| - $R^2$ | continuous dependent variable |

- **Several correlated models can be considered together**

## Impact of model quality analysis: simulated data

- 10 000 customers
- 3 decision options
- Decision option properties:

  response propensity

  expected revenue

  cost of solicitation

  cost of selling
- Metrics:

  number of solicitations

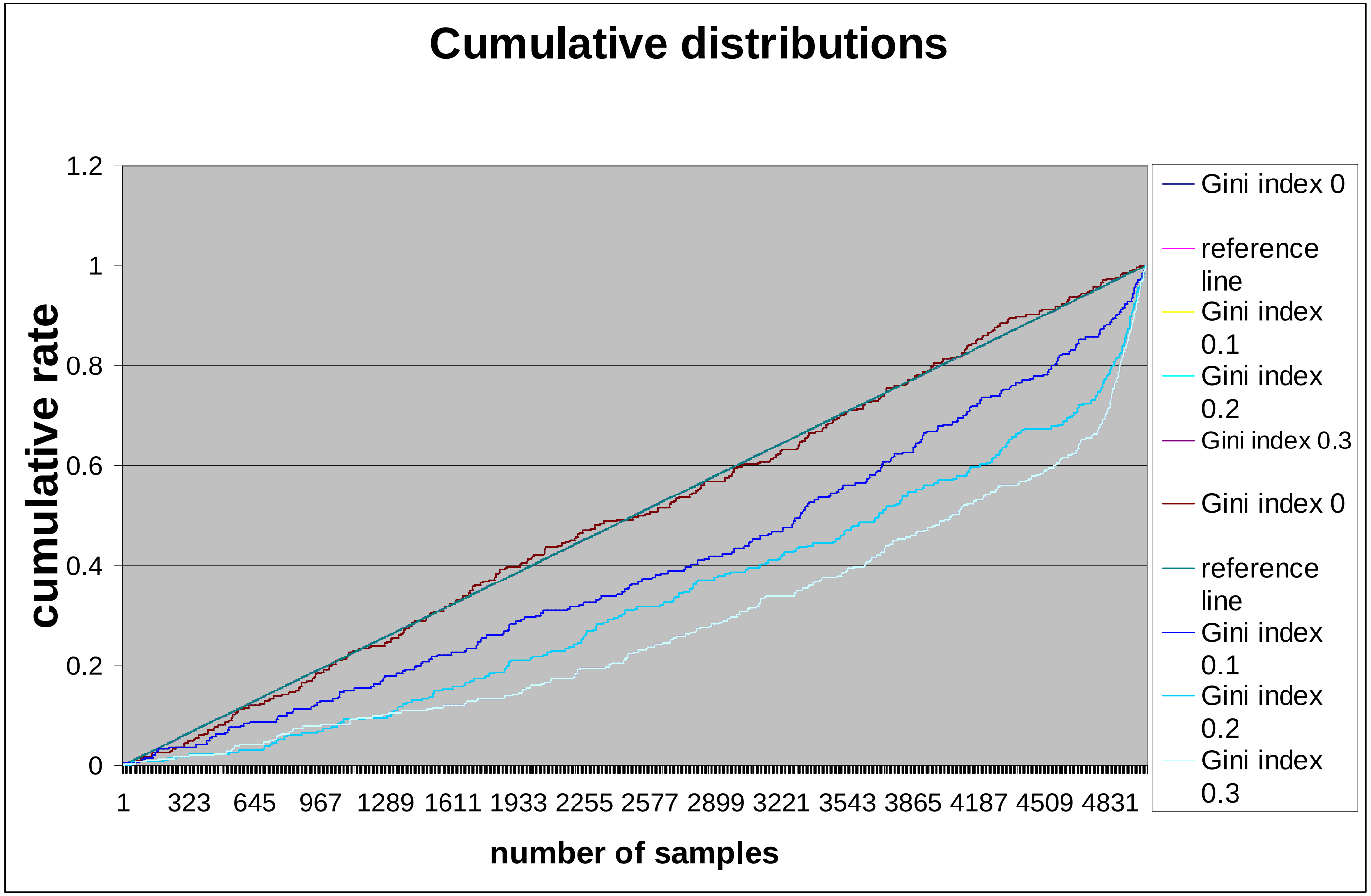
  number of sales

  revenue

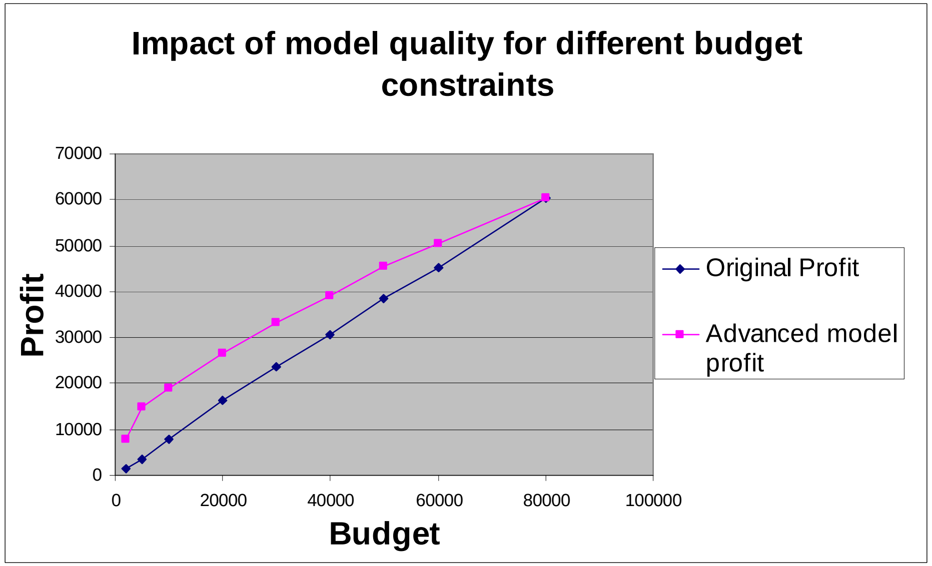
  profit

  cost
- Utility functions: profit, number of sales, cost
- Constraints: number of solicitation, number of sales, cost, profit per customer

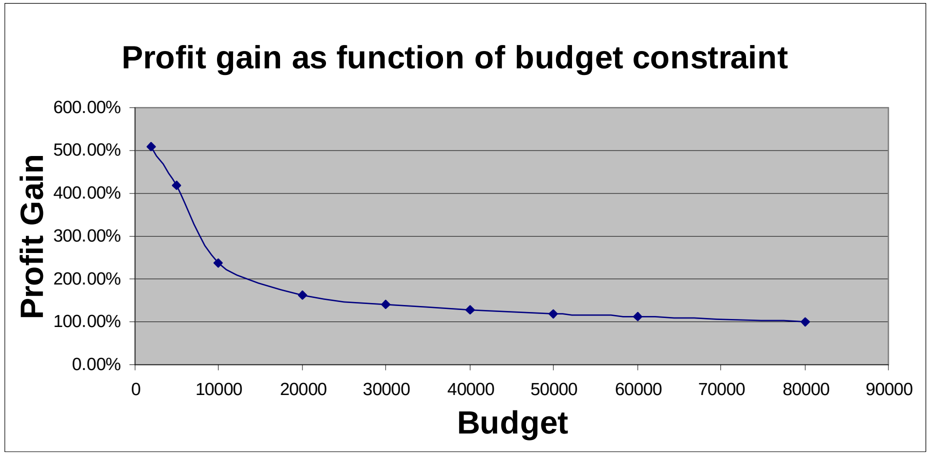
# Impact of model quality analysis: cumulative distributions for different values of model quality measure

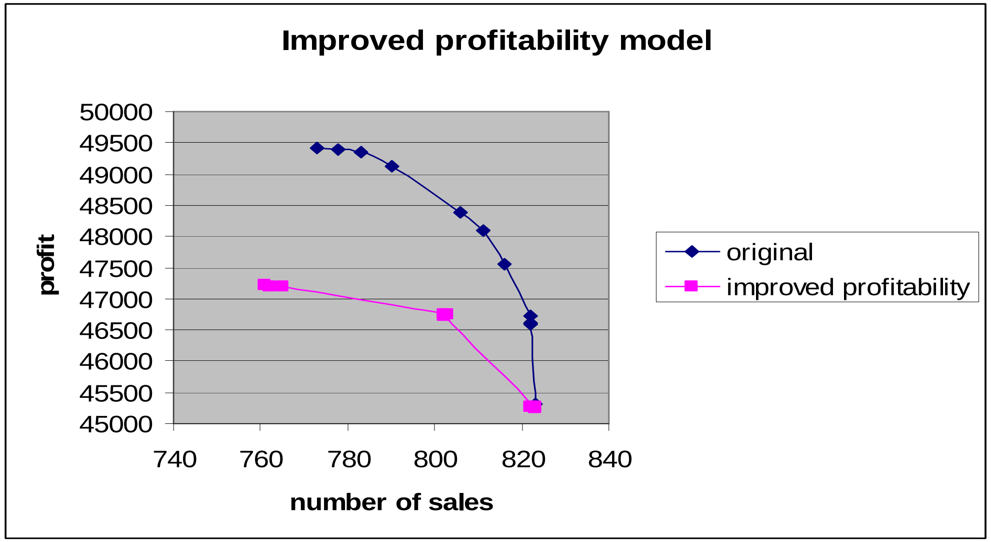


# Impact of model quality analysis: role of constraints



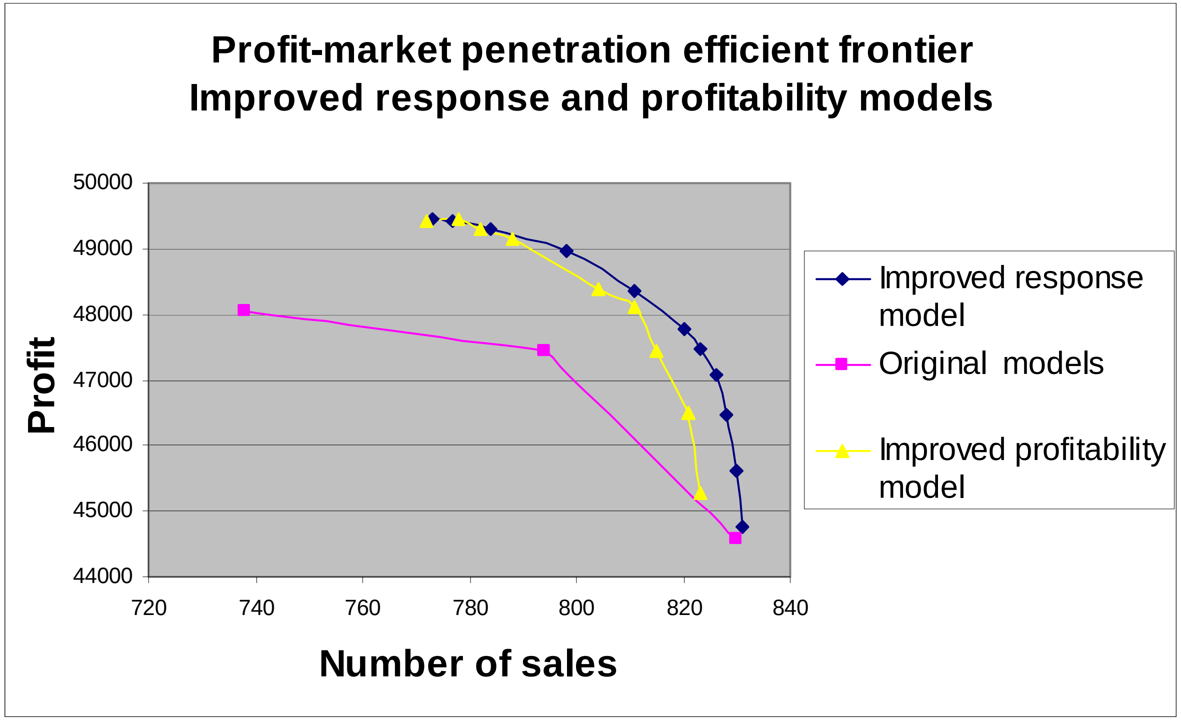

**In this example we see that when resources become more restricted the gain due to increased model quality becomes more significant. Effect of model quality is especially significant for tightly constrained optimisation cases**



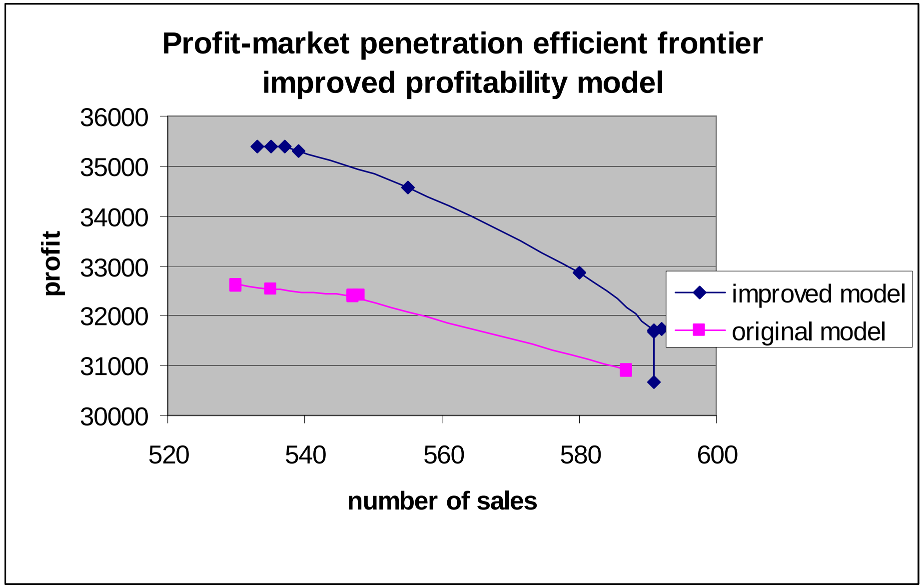



# Impact of model quality analysis: choice of objective function
# Results for response and profitability models

This example shows that for given optimisation task not all models are equally important. Response model improvement has more noticeable effect in all variety of profit- market penetration trade-offs for this particular case



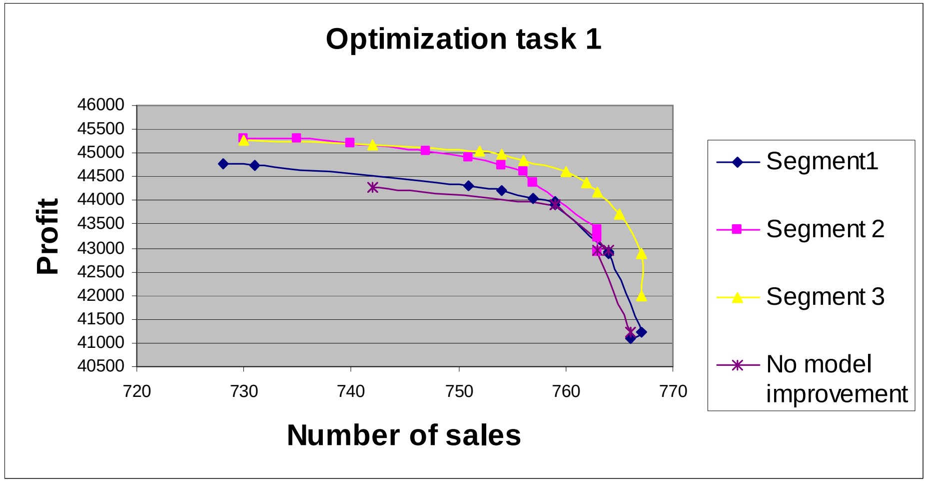
# Impact of model quality analysis: taking correlation into account

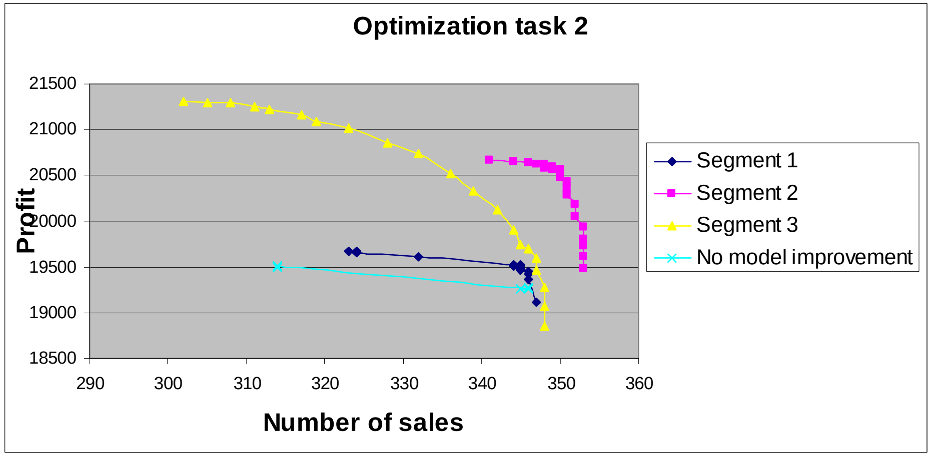



The following example shows the expected increase in both profit and number of sale for 10% improved profitability and response models with 0.5 negative correlation.

## Impact of model quality analysis:
## Results for different data segments: comparison of two optimisation tasks

In this example we analyze the relative importance of model improvement for different data segments. Average response rate for Segment 1 is high, for Segment 2 is intermediate, and Segmen3 is small. Average segment profitability is also different between segments




The results show that depending on optimisation task (set of constraints) and choice of compromise between objective functions either Segment 1 or Segment 2 can be performance critical and may be chosen for model update

Experian™

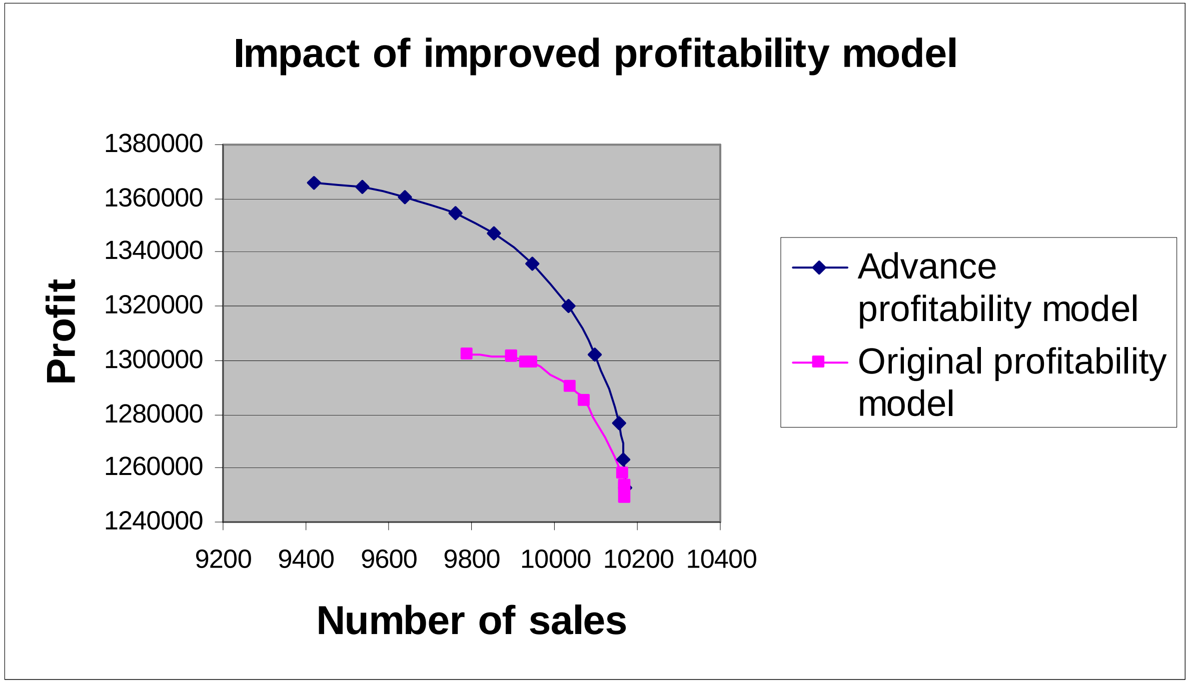
# Impact of model quality analysis: a major bank case study

- **Decision options**: 9 types of acquisition loans, 12 types of retention loans, 2 types of classic credit cards, 2 types of gold cards, 6 type of platinum cards
- **Metrics and objective functions:** profit, market penetration, cost
- **Models**

  response propensity models: available for each decision option (even with taking cannibalization effect into account)

  profitability models: only average values for each decision option
- **Conditions:** maximum one loan and one credit card promotion can be selected  for each customer

# Impact of model quality analysis: a major bank case study




Effect of improved model quality for the data case described above. No correlation between profitability and response model

# Impact of model quality analysis: a major bank case study Correlation and adverse selection effect

No correlation

| Objective function | Profit | Advanced model profit | Number of sales |
|---|---|---|---|
| Original profitability model | 1289037 | 1289866 | 9566 |
| Advanced profitability model | 1241231 | 1363900 | 9402 |

0.3 negative profit/response propensity correlation

| Objective function | Profit | Advanced model profit | Number of sales |
|---|---|---|---|
| Original profitability model | 1289037 | 942903 | 9566 |
| Advanced profitability model | 1221178 | 1016520 | 9192 |

# Conclusion

- Only precise record level optimisation can utilize the advanced model quality to full extend
- Analyzing impact of model quality will allow practitioners to make non-trivial predictions of how model enhancement affects the results for different optimisation tasks
- Impact of model quality can be strongly influenced by choice of objective function and constraints
- Effect of model quality is especially significant when multiple constraints are applied and profit margin is small, which is typical in current economic environment
- The effect of model quality is not necessarily uniform across different segments.  The analysis allows to identify which segment of the data is performance critical for a given optimisation task and concentrate on improving of model quality for this segment alone
- In the case when metrics values are correlated, the approach allows to evaluate the distortion effect of  using poor quality models
- This method can be used as a tool for cost benefit analysis of model improvement efforts and help to create a roadmap for building, improving, and updating predictive models