

QUALCO

Portfolio Analytics Department

**Collections Data Capitalization
on Point-in-Time PD Models for
Retail Portfolios**



**Credit Scoring and Credit Control XIII Conference
Edinburgh, August 28-30, 2013**

➤ Concept

➤ Case Study

➤ Next Steps

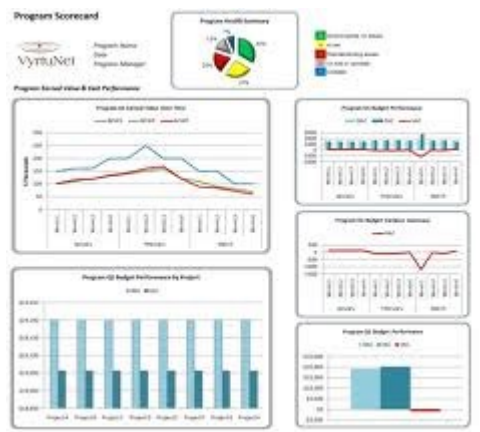
Unsecured Consumer Loans Portfolio

Input

6.8	3.2	2.1	2.1	3.0	3.4	6.2	14.3	13.0
7.5	4.0	2.5	3.0	2.4	4.0	6.1	11.5	10.5
8.5	4.9	2.6	1.6	2.4	2.0	5.3	10.0	9.6
9.5	4.3	2.0	1.6	1.9	2.0	4.9	8.2	8.7
10.5	3.3	2.8	1.8	1.5	2.0	4.9	7.6	7.4
11.5	2.9	2.9	1.9	1.4	2.1	3.9	6.3	6.8
12.5	3.0	2.8	1.8	1.5	1.8	3.3	4.9	5.3
13.5	3.2	2.6	1.7	1.5	1.8	3.6	4.0	4.6
14.5	3.7	2.7	1.6	1.6	1.8	2.3	3.6	4.0
15.5	3.0	3.0	1.6	2.0	1.8	2.2	3.4	3.5
16.5	3.0	3.4	1.8	1.4	1.6	2.2	3.2	3.0
17.5	3.5	3.5	2.2	1.6	1.7	2.4	3.0	2.4
18.5	3.2	3.4	2.0	2.0	2.1	2.7	2.5	1.8
19.5	2.5	2.9	3.1	2.8	2.8	2.7	1.9	1.2
20.5	1.8	2.3	3.0	3.4	3.2	2.2	1.3	0.8
21.5	1.1	1.7	2.5	3.5	3.1	1.6	0.8	0.5
22.5	0.6	1.2	2.0	3.3	2.7	0.9	0.5	0.4
23.5	0.4	0.8	1.6	2.8	2.2	0.6	0.4	0.3
24.5	0.2	0.5	1.1	2.3	1.7	0.4	0.3	0.2
25.5	0.1	0.3	0.8	1.8	1.5	0.3	0.2	0.2

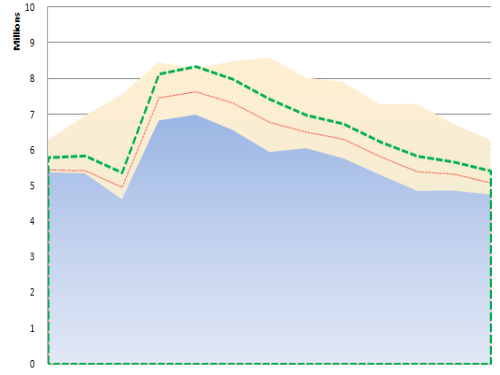
+
Add Collections data

PD/LGD Models

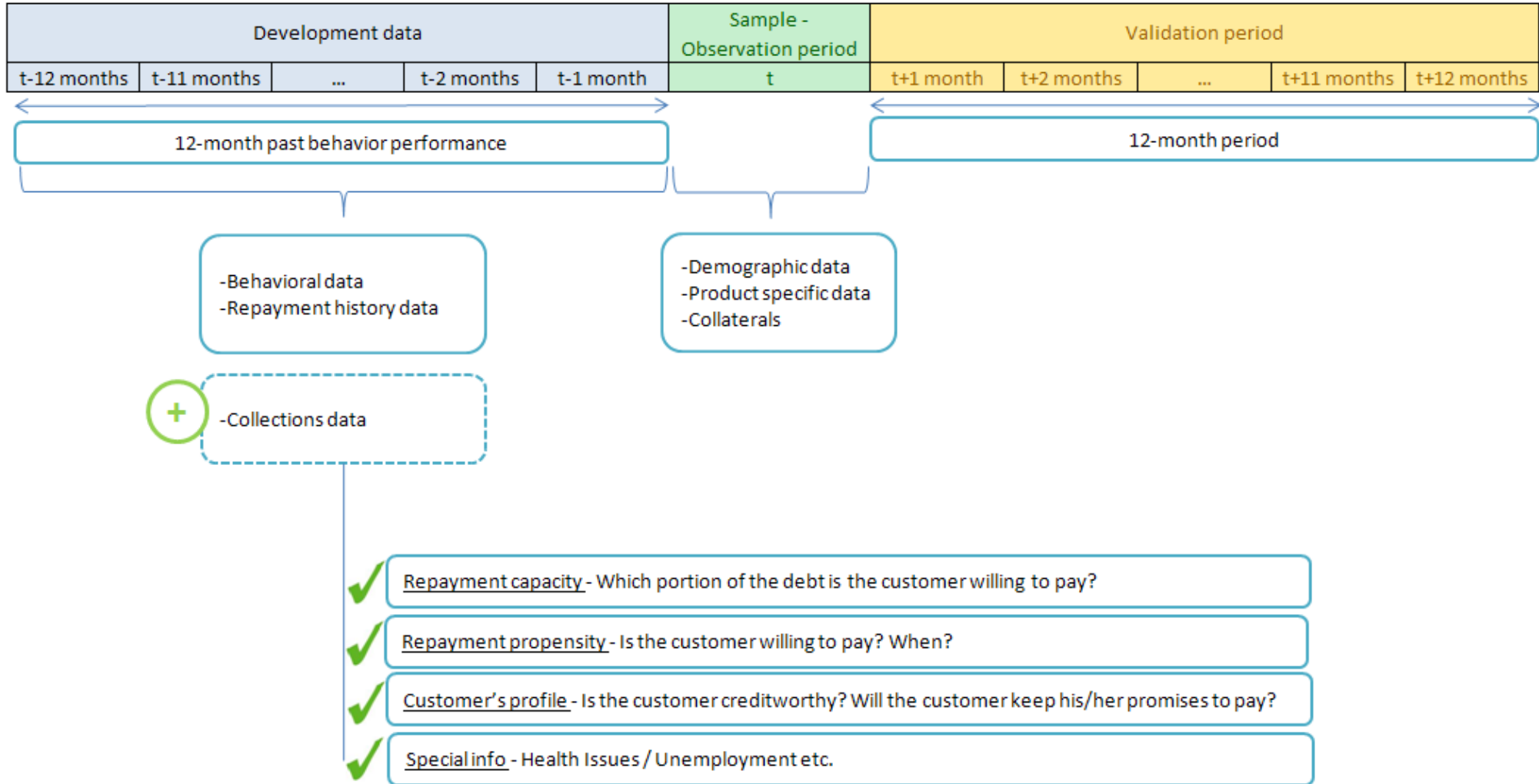


PD Model (Logistic Regression)
Measure uplift

Expected Losses



Compare Expected Losses



“Qualitative” data

Offers a qualitative aspect of customer’s profile – captures information untapped from “traditional” sources.

Time advance

Offers a competitive advantage; a deterioration of financial situation is being captured before delinquency increase.

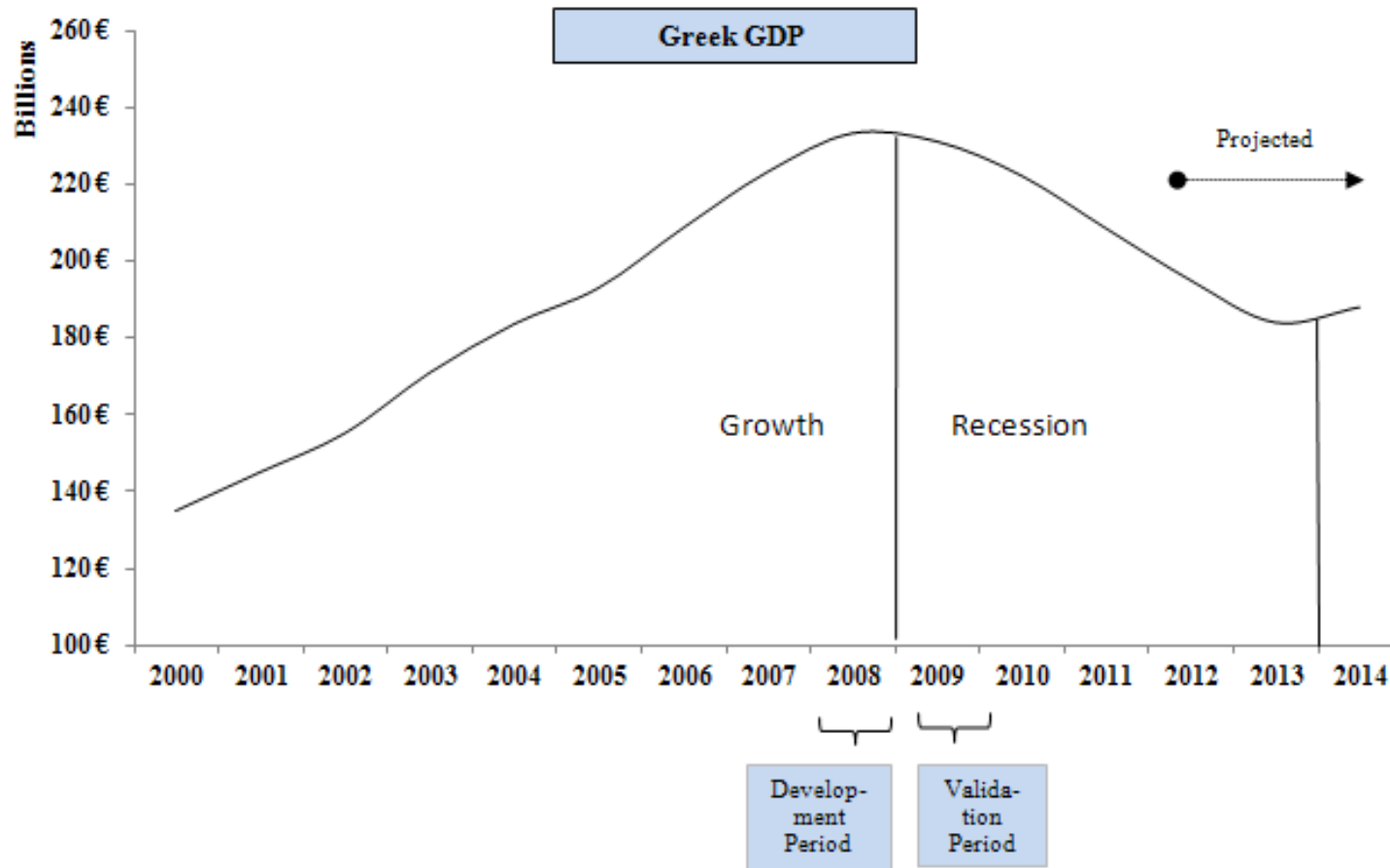
Extra Dimension

Eventually it becomes an additional dimension in customer’s risk assessment

- Product
- Delinquency
- Balance
- Collections’ Profile**



- Concept
- Case Study
- Next Steps



- SE European bank, IRB compliant, operates in the Greek market.
- Unsecured consumer loan portfolio, 1-59 dpd.

Collections Data

Data derived from the collection system of the institution and thus represents the information associated with the collection process

Process

Collection model was trained, developed and validated using exactly the same definitions as the Initial PD model so as to have consistent and comparable results.

Capitalize Data

Transaction data was used to calculate aggregate variables such as kept promise ratio, contacts to activities ratio, right person contacts and a significant number of other key performance indicators (over 250 variables were tested) during a performance period of 12 months.

Binning

$$IV = \sum_{i=1}^j 100 * \left[\frac{G_i}{Goods} - \frac{B_i}{Bads} \right] * \ln \left[\frac{G_i / Goods}{B_i / Bads} \right]$$

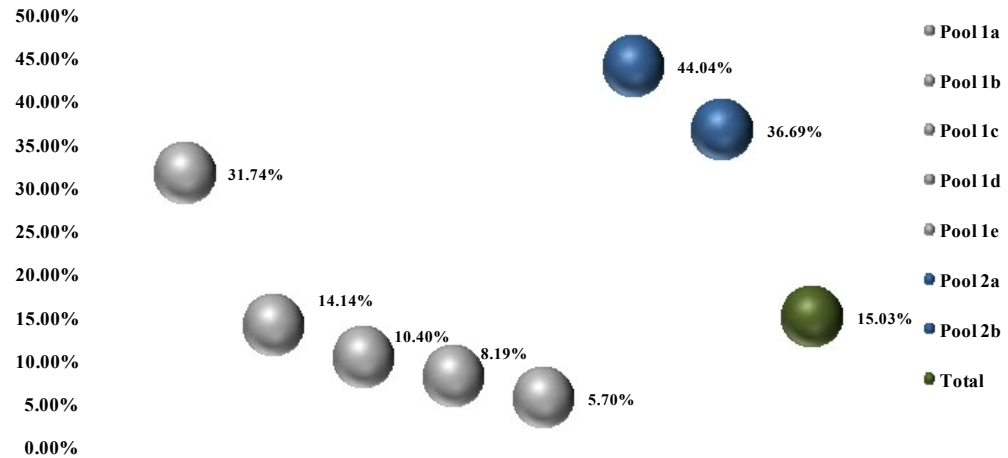
Discrimination Power

The final collections model is discriminating satisfactorily - Gini Index = **46.4%**
 Initial PD Model is slightly more powerful - Gini Index = **48.4%**

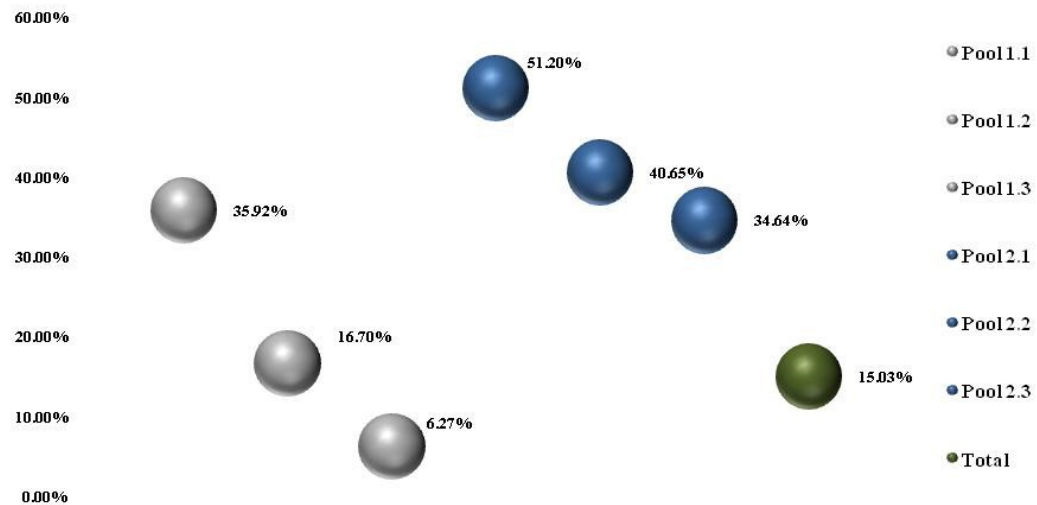
I. Enhanced Pools - Step 1

■ The portfolio was split into two different sets of segments (based on days of delinquency) and pools based on probability of default ratios of each model as indicated by the development sample.

PD per Segment Initial Pd model
Based on Development Period



PD per Segment Collection model
Based on Development Period



I. Enhanced Pools – Step 2

■ The next step was to combine the output of the two models. Pools as indicated by the Collection Model were used as sub-segments in order to further split Initial PD models pools.

Segment	1st Dimension	2nd Dimension	Enhanced Pool
	Behavioral Pool	Collections Pool	
Segment 1 (1-29 days past due)	Pool 1a	Pool 1.1	Pool 1a-1.1
		Pool 1.2	Pool 1a-1.2
		Pool 1.3	Pool 1a-1.3
	Pool 1b	Pool 1.1	Pool 1b-1.1
		Pool 1.2	Pool 1b-1.2
		Pool 1.3	Pool 1b-1.3
	Pool 1c	Pool 1.1	Pool 1c-1.1
		Pool 1.2	Pool 1c-1.2
		Pool 1.3	Pool 1c-1.3
	Pool 1d	Pool 1.1	Pool 1d-1.1
		Pool 1.2	Pool 1d-1.2
		Pool 1.3	Pool 1d-1.3
	Pool 1e	Pool 1.1	Pool 1e-1.1
		Pool 1.2	Pool 1e-1.2
		Pool 1.3	Pool 1e-1.3
Segment 2 (30-59 days past due)	Pool 2a	Pool 2.1	Pool 2a-2.1
		Pool 2.2	Pool 2a-2.2
		Pool 2.3	Pool 2a-2.3
	Pool 2b	Pool 2.1	Pool 2b-2.1
		Pool 2.2	Pool 2b-2.2
		Pool 2.3	Pool 2b-2.3

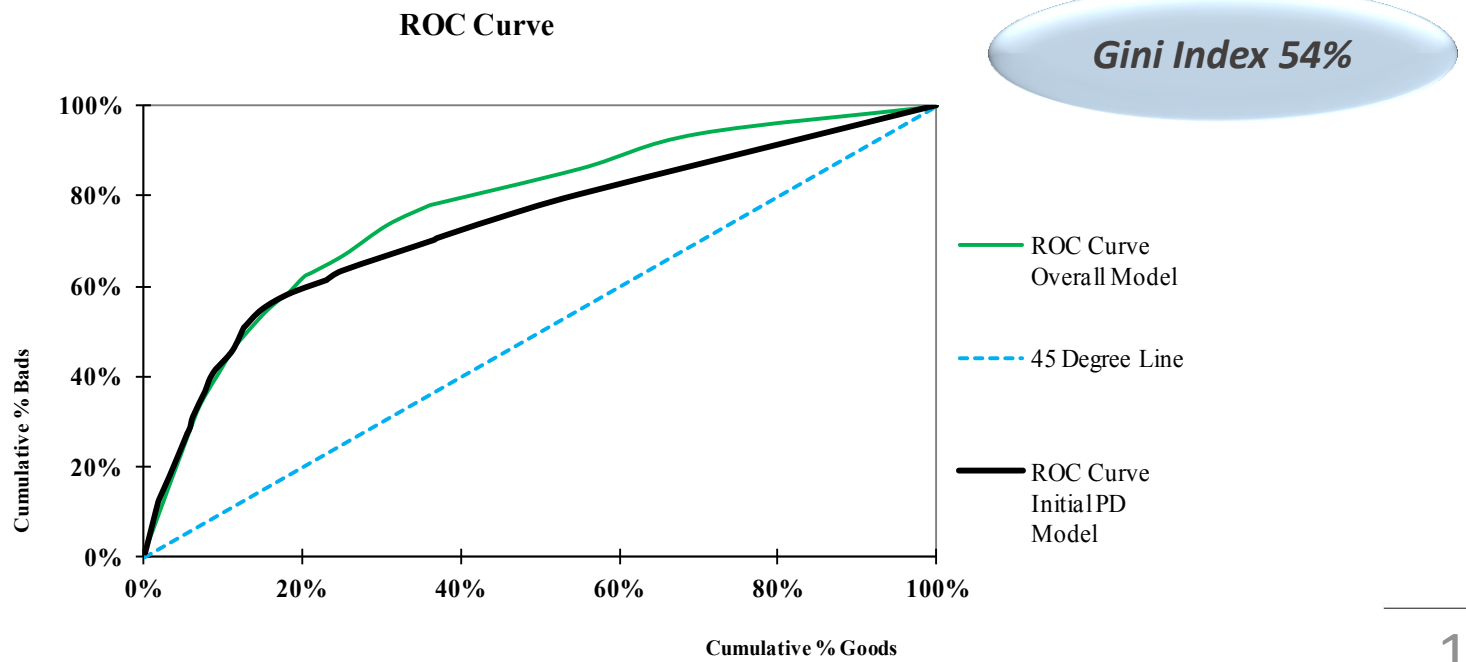
Overall Model Training

A logistic regression was performed in order to develop an Overall model, using as input variables the Initial PD model and the Collection model.

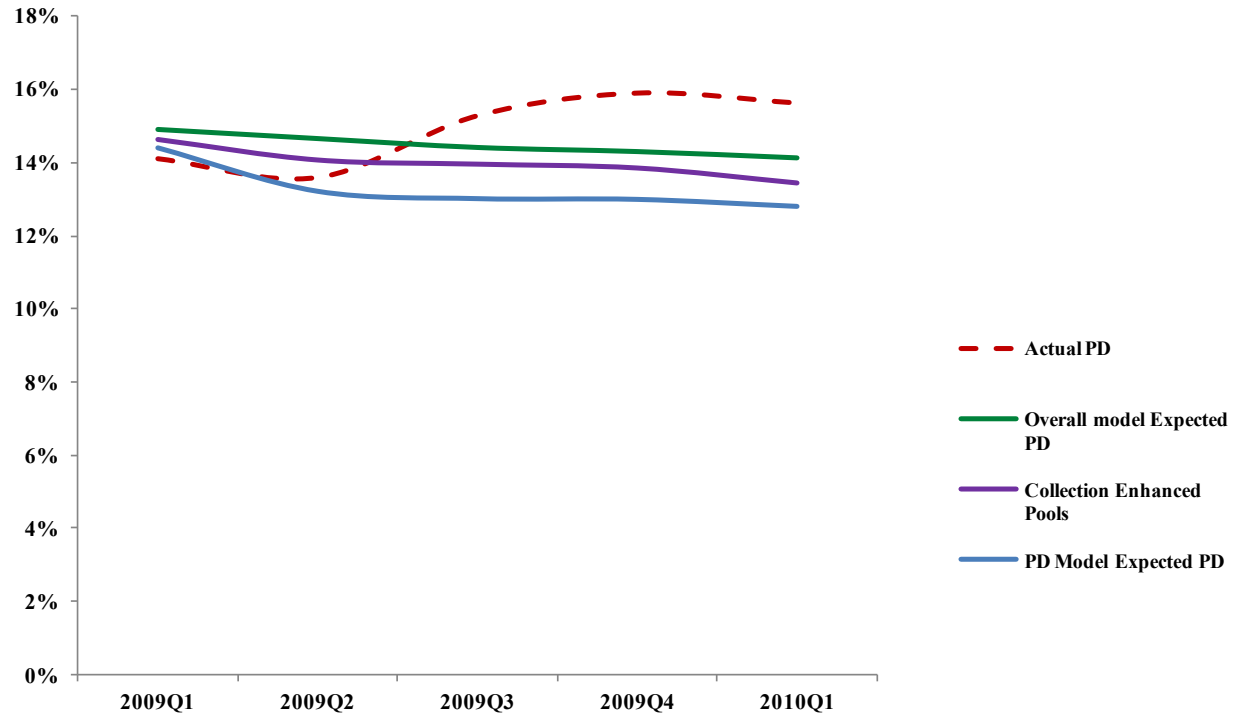
Consistency

The overall model was trained in the same sample data as both the Initial PD model and the Collection model in order to maintain consistency and to achieve comparable results.

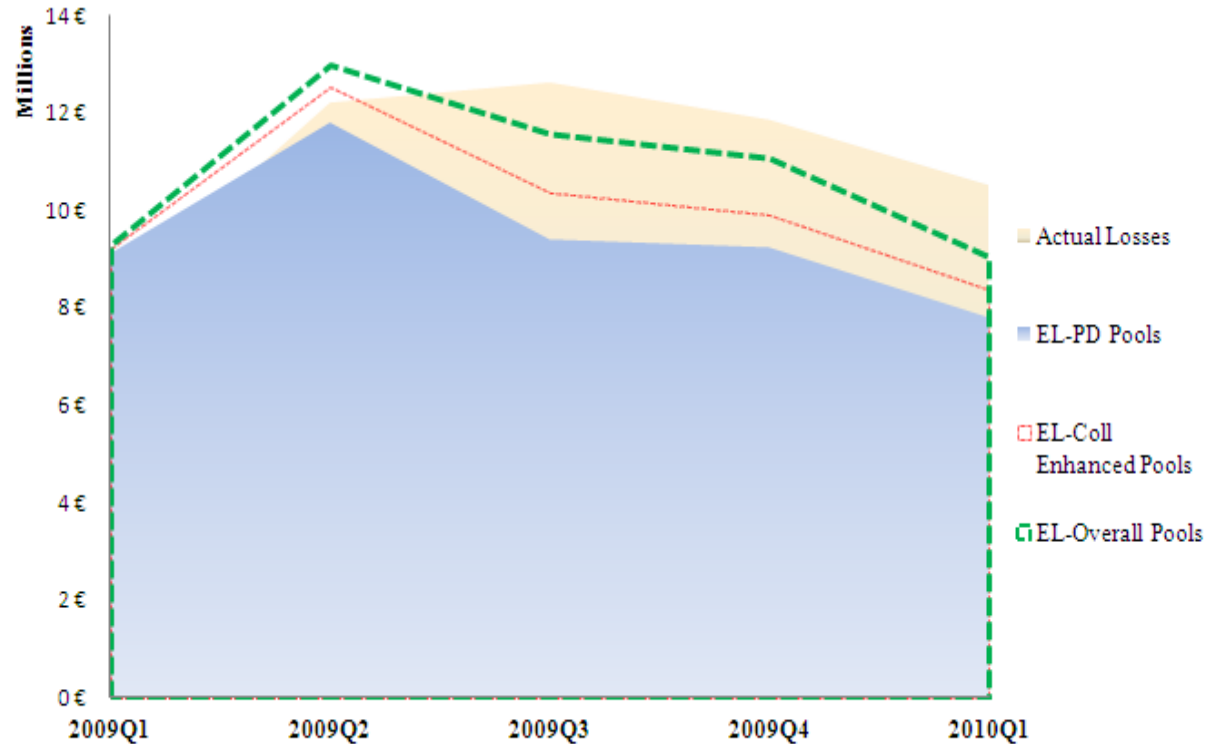
Discrimination Power



- Actual PD is increasing as we move towards recession period.
- Initial PD model (behavioral PD) is following a completely different trend.
- Expected PD derived from collections enhanced pools is closer to the actual PD; still is affected by the behavioral model's trend.
- Overall model's expected PD is far better than anything else in approaching actual PD.



- Common LGD period (18-month post default)
- Expected losses were calculated in order to be compared with the above-mentioned PD sets, having the LGD part of the formula in common.
- PD and LGD models developed before recession fail to approach the losses that occur during the recession period.
- On the other hand, collections information succeeds in bridging the gap between EL and Actual Losses.



- Concept
- Case Study
- Next Steps



Repeat the exercise with different portfolios
(mortgage / small business)



Repeat the exercise in a different period in the
economic cycle



Utilize collections data in LGD models

Thank you

Q&As