UNIVERSITY OF EDINBURGH
Business School

CRC Credit Research Centre

THE COMPARATIVE ANALYSIS OF PREDICTIVE MODELS FOR CREDIT LIMIT UTILIZATION RATE

PROFESSOR JONATHAN CROOK
DENYS OSIPENKO

CRCCXIV, 26-28 August 2015, Edinburgh

Content

- Objectives
- The utilization rate definitions and usage
- General model description
- One-stage model comparative analysis:
  - OLS
  - Beta-regression
  - Beta transformation plus GLM
  - Fractional regression (quasi-likelihood)
  - Weighted logistic regression with binary transformation
- Two-stage model comparative analysis:
  - Probability of use
  - Direct estimation
- Conclusions

Objectives of investigation

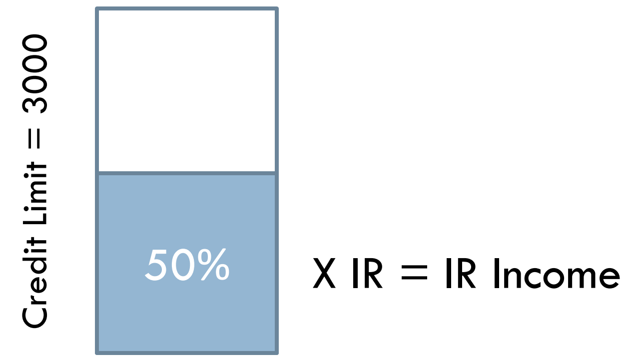
- We perform a comparative analysis of a set of methods for the credit limit utilization rate (usage proportion) prediction:

i) five direct estimation techniques such as:
   - ordinary linear regression,
   - beta regression,
   - beta transformation plus general linear models (GLM),
   - fractional regression (quasi-likelihood estimation),
   - weighted logistic regression for binary transformed data,

ii) two-stage models: probability of use PLUS the credit limit utilization rate direct estimation

Credit line Income Prediction

- For interest income
  - Utilization = Balance / Credit Limit
  - IR_Income = Utilization Rate x Limit x IR
- For non-interest income (POS, ATM, Interchange etc.)
  - POS Income = TR Debit_POS x POS_fees_rate
  - Interchange = TR Debit_POS x Interchange_fees_rate
  - Cash Withdrawal Income = TR_Debit_ATM x ATM_fees_rate



Interest Income from Balance:
1500 UAH X 36% /12 = 45 GBP

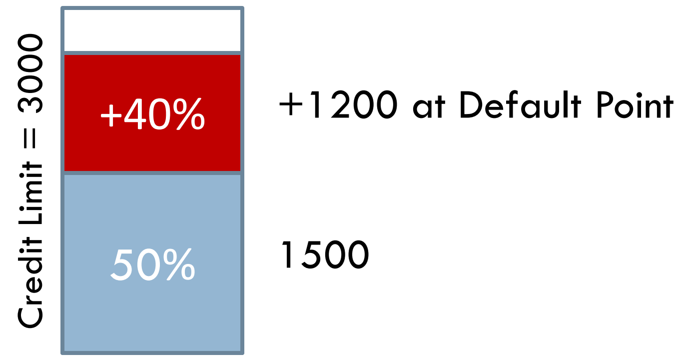
Monthly transactions:
1000 UAH POS X 2% = 20 GBP
500 UAH ATM X 2,5% = 12,5 GBP

Total Non_Interest Income = 32,5 GBP
Total Income = 45+32,5 = 77,5 GBP

Exposure at Default estimation

Expected Loss

EL = PD x LGD x EAD



Exposure at Default for credit card:

$$EaD = L \cdot UR + \left(L\left(1 - UR\right)\right) \cdot CF$$

CF = 1200/1500 = 80%

CF (conversion factor) – the percent (share) of the additional usage of remaining credit line at the default point.
The credit conversion factor (CCF) converts the amount of a free credit line and other off-balance-sheet transactions (with the exception of derivatives) to an EAD (exposure at default) amount.
L – credit limit

Some investigations of EaD - Jacobs (2008), Qi (2009)

References

- RamonaK.Z.Heck (1987). Differences in Utilisation Behaviour Among Types of Credit Cards. The Service Industries Journal Volume 7, Issue 1, 1987.
- Bellotti T. and Crook J.(2009).Loss Given Default models for UK retail credit cards. CRC working paper 09/1.
- Bellotti, T. and Crook, J. (2012). Loss given default models incorporating macroeconomic variables for credit cards. International Journal of Forecasting, 28, 171-182.
- Arsova, M. Haralampieva, T. Tsvetanova (2011). Comparison of regression models for LGD estimation. Credit Scoring and Credit Control XII 2011, Edinburgh.
- Stefan Stoyanov (2009). Application LGD Model Development. A Case Study for a Leading CEE Bank. Credit Scoring and Credit Control XI Conference, Edinburgh, 26th-28th of August 2009
- Steven Xizogang Wang (2001). Maximum weighted likelihood estimation. The University of British Columbia. A part of thesis published, June, 2001.
- Sumit Agarwal Brent W. Ambrose Chunlin Liu (2006). Credit Lines and Credit Utilization. Journal of Money, Credit, and Banking, Vol. 38, No. 1 (February 2006).
- Papke, L. E. and Wooldridge, J. M. (1996), “Econometric Methods For Fractional Response Variables With an Application to 401(K) Plan Participation Rates”, Journal of Applied Econometrics, vol. 11, 619-632
- Ferrari, S. L. P. and Cribari-Neto, F. (2004), “Beta Regression for Modelling Rates and Proportions”, Journal of Applied Statistics, 31, 799-815
- Anthony Van Berkel, Bank of Montreal and Naeem Siddiqi (2012). Building Loss Given Default Scorecard Using Weight of Evidence Bins in SAS®Enterprise Miner™. SAS Institute Inc. Paper 141-2012
- Xiao Yao, Jonathan Crook, Galina Andreeva. (2014). Modeling Loss Given Default in SAS/STAT®. SAS Forum 2014. Paper 1593-2014.
- Osipenko D & Crook J (2015). The Comparative Analysis of Predictive Models for Credit Limit Utilization Rate with SAS/STAT®. SAS Institute, April 2015. Paper 3328-2015.

Utilization Rate modelling

- Utilization rate = Balance / Credit Limit
- Regression equation: utilization rate depends from behavioural, application, macroeconomic characteristics, and also from utilization rate with time lag

$$UT_{it} = \phi_1 UT_{i(t-1)} + \ldots + \phi_T UT_{i(t-T)} + \sum_k^K \beta_b \cdot B_{bi,t-1} + \sum_l^L \alpha_a \cdot A_{ai} + \sum_m^M \gamma_1 M_{m,t-1}$$

φ, α, β, γ – regression coefficients (slopes)

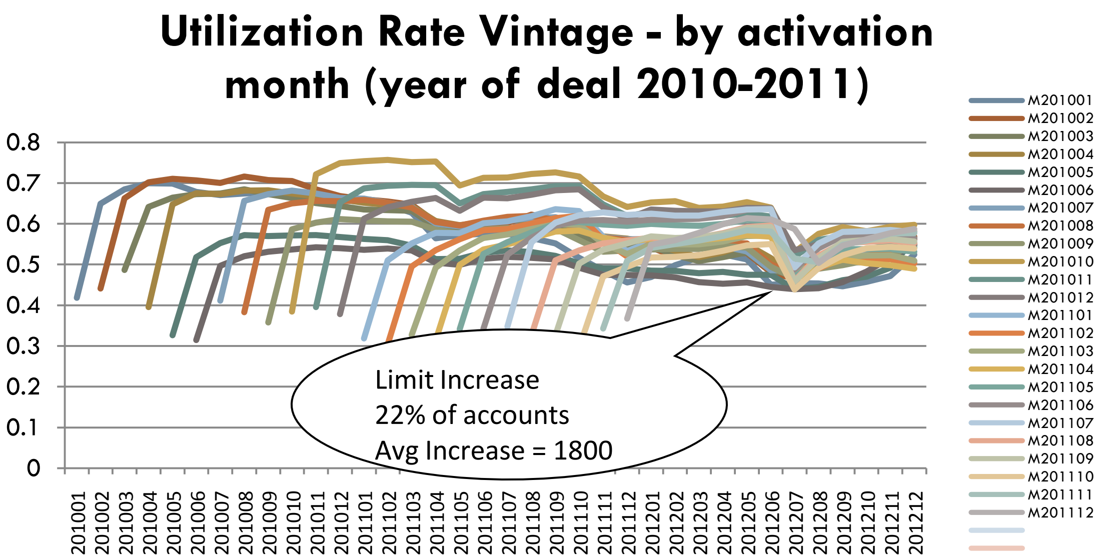
B *it* **–** vector of behavioural factors *b* (for example, average balance to maximum balance, maximum debit turnover to limit etc for period *t* observation *i*)

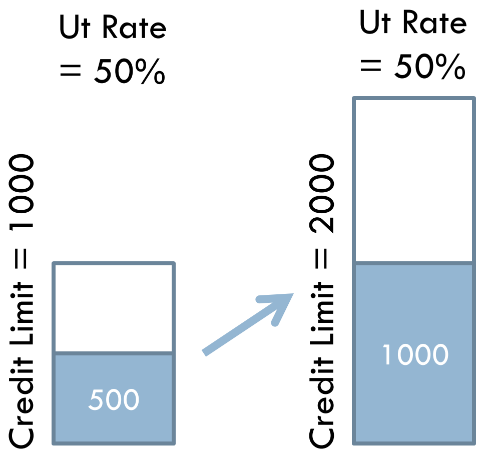
A *i* – vector of application factors - client's demographic, financial and product characteristics like age, education, income etc. for observation *i*

M *t* – vector of macroeconomic factors (GDP, FX, Unemployment rate changes, etc.)

UT – utilization rate

Why the utilization rate, not the outstanding balance



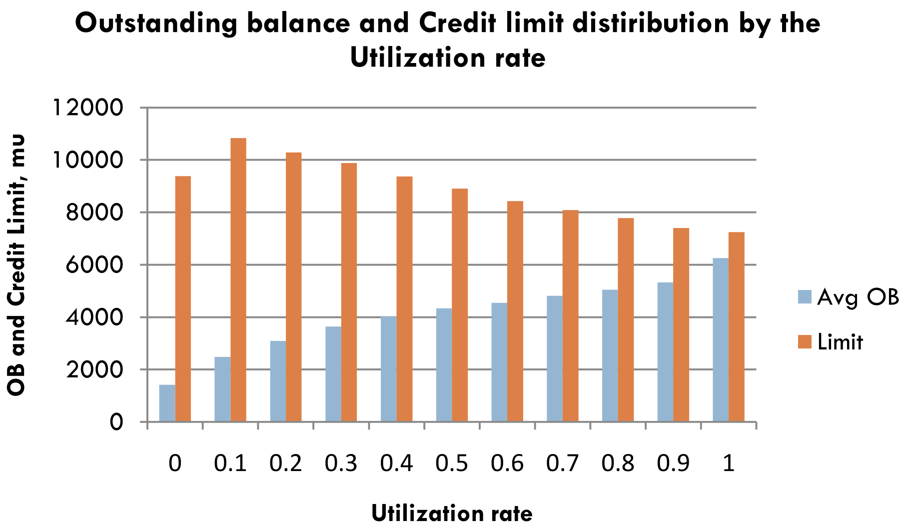


Limit increase process in June has shown the utilization rate drop, but it has stabilized **almost at the same** level in two months.

Assumption: Customer behaviour characteristics and customer profile rather have impact on the utilization rate than on the outstanding balance.

Credit limit depends on credit policy rules and sets up particularly according to the customer risk profile. The same behavioural customer segments have various outstanding balances correlated particularly with the credit limit. Thus customer segment does not has typical outstanding balance, but typical utilization rate as proportion of the credit limit.

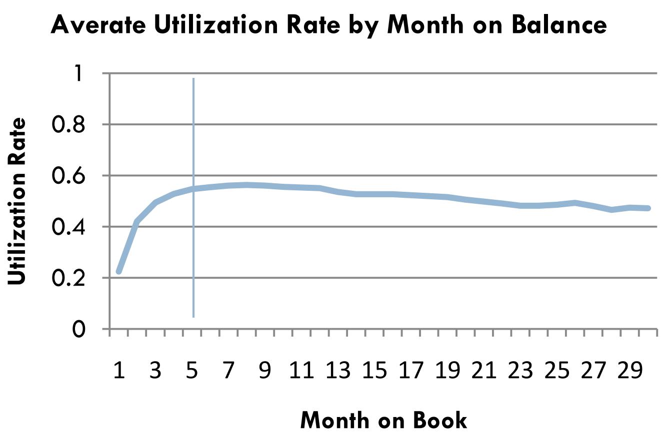
Model segments empirical explanation



The utilization rate decrease can be caused by the credit limit increase process. However, the customer behaviour can be changed because of the limit changes. This is the reason why it is reasonable to split the model up two segment:

i) credits with unchanged limits and
ii) credits with the limits which have been changed.

Because of the inconsistencies in the behavioural characteristics calculation (lack of history) and the differences in the utilization rate dynamic at the early and late credit history stages it is rational to allocate the separate model for the low MOB period. In our case two periods have been chosen: MOB from 1 to 5 and MOB more than 6.



Model segments

The data sample has given maximum 30 month period for investigation. First 24 months are used as an observation period for behavioural characteristics calculation. Available months on balance after 25th month are used as a performance period only.

<table>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13 …</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td></tr>
<tr><td colspan="5" rowspan="3">Model APP: MOB 1-5</td><td colspan="9">Model BEH NL: MOB 6+ and Limit NO Change</td><td colspan="6" rowspan="2">Performance</td></tr>
<tr><td colspan="9">Model BEH CL: MOB 6+ and Limit Changed</td></tr>
<tr><td colspan="15">Performance</td></tr>
</table>

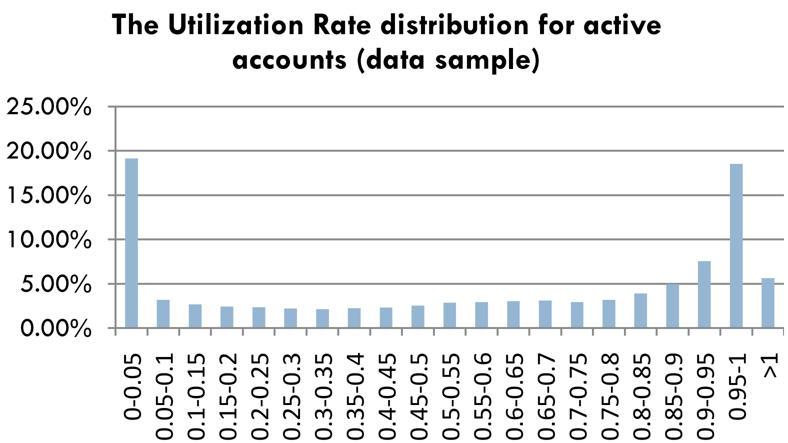
For the first 5 months on balance the model called the Application model (Model APP) is applied.

This model contains application and short term behavioural characteristics.

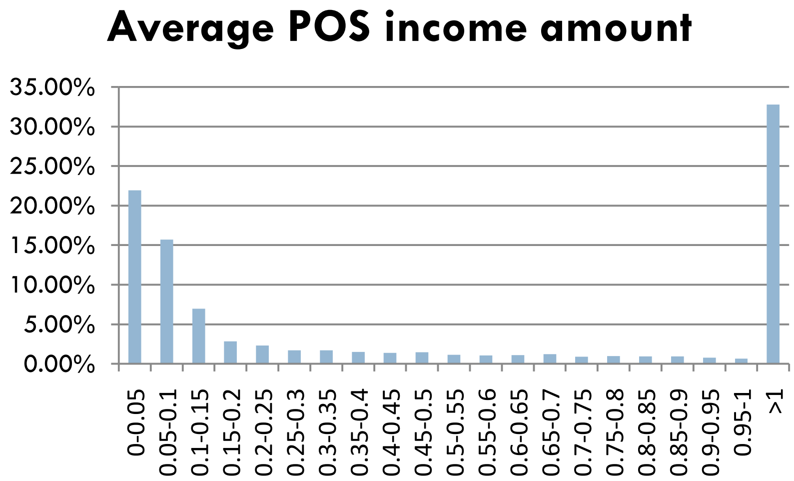
Two long term behavioural models as non-changed limit (BEH NL) and changed limit (BEH CL) are applied for the next 18 months (from 6th to 24). Also this period from 6th to 24th month on balance is used as a performance window for the development and validation with appropriate lag for any previous observation window.

For example, the loan activated in July 2011 has 6 month history from July till Dec and 6 month performance period from Jan to June 2012.

Utilization rate and Income distributions

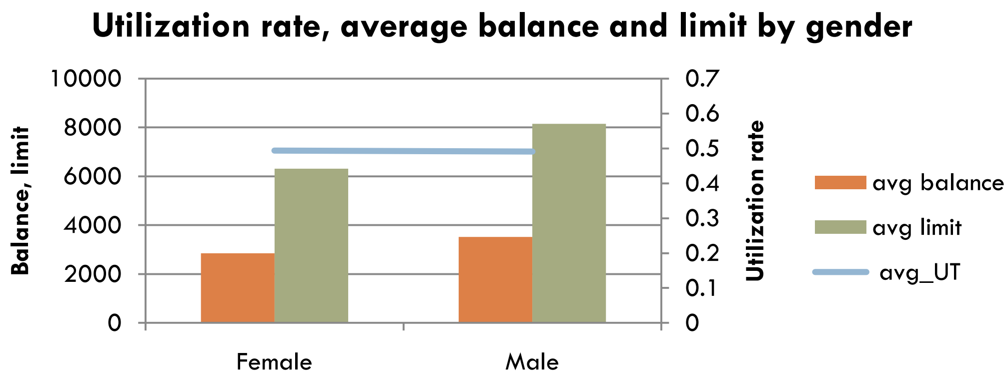


Utilization rate density may have an U-shape distribution,
can be approximated, as option, with beta-distribution.
Values >1 caused by overlimit, are temporary and replaced by 1.

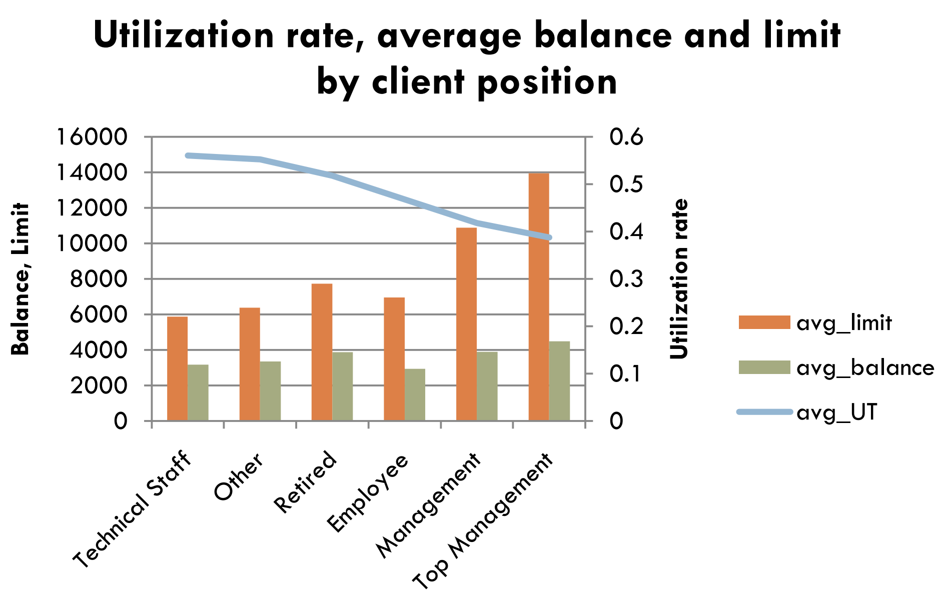


POS income (interchange fees) may have exponential distribution
It’s necessary to filter a lot of insufficient amounts and enormous outliers

Utilization rate, balance and limit for gender and age distributions

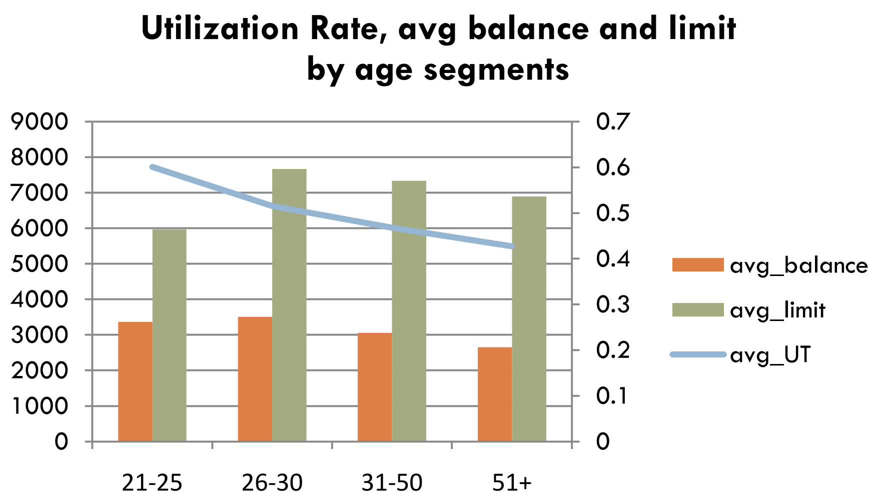


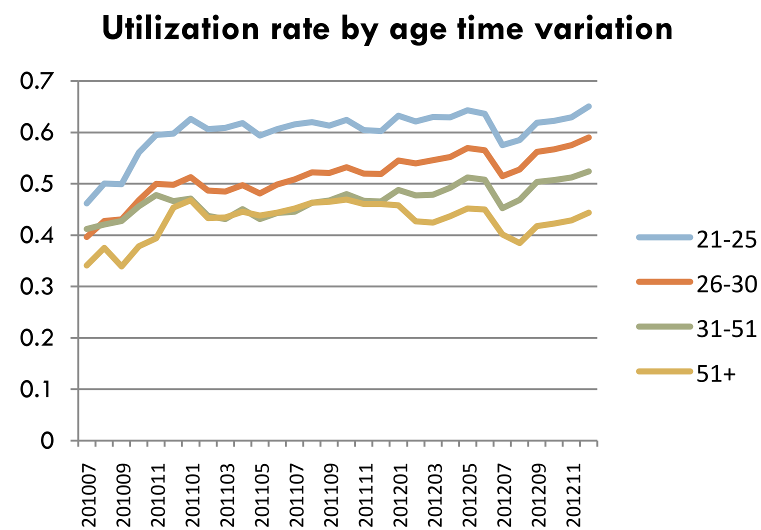
The utilization rate for gender is not differed significantly, but the average balance for male is higher than for female most likely because of higher limits.

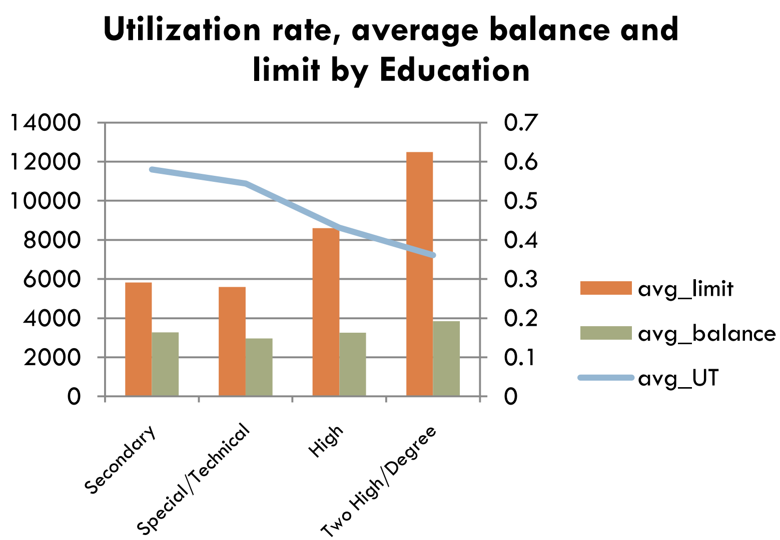


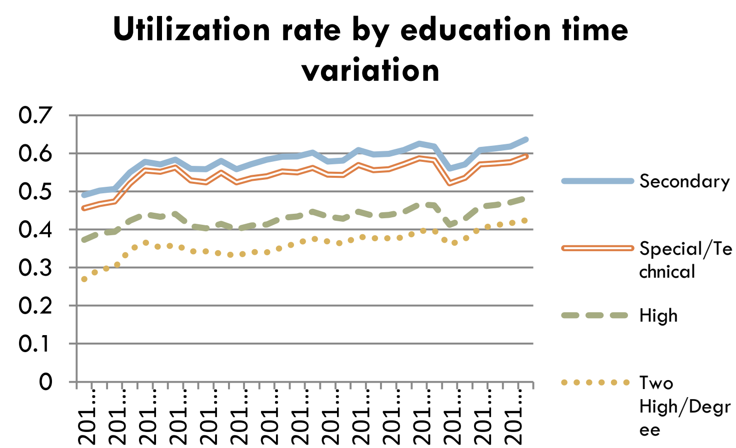
Top managers have the highest limits, but the lowest utilization rate in compare with other positions. This means that the outstanding balance is not so different as the credit limit.

Utilization rate, balance and limit Age and Education distributions









Regression methods (1/2)

Linear regression - OLS

$$y_i = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

Beta regression

$$f_X(x) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Because the outcome is in the range between 0 and 1 the logistic transformation is used to find the dependences between predictors x(a) and rearessor

$$Beta(y, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \qquad \mu(a) = L(\mathbf{x}(a)'\beta) = \frac{e^{\mathbf{x}(a)'\beta}}{1+e^{\mathbf{x}(a)'\beta}}$$

Beta-transformation plus OLS

The algorithm uses the beta distribution to transform the original target.

1 - to find the beta-distribution coefficients (alpha and beta) - fit the sample distribution using the non-linear regression procedures.

2 - replace real target variable by the ideal beta-distributed.

3 -find appropriate normal distributed value.

4 - run OLS or Generalized Linear Mixed Model to find regression coefficients.

Regression methods (2/2)

Fractional logit transformation (quasi-likelihood)

The utilization rate has is bounded between 0 and 1. The Bernoulli log-likelihood function is given by

$$l_i(\mathbf{b}) \equiv y_i \log[G(\mathbf{x}_i\mathbf{b})] + (1 - y_i)\log[1 - G(\mathbf{x}_i\mathbf{b})]$$

$$T_{UT} = \log(UT) - \log(1 - UT)$$ - Transformation function

Weighted Logistic regression with binary transformation

Utilization Rate	Binary recovery – target	Weight
1	1	1
0	0	1
R, 0<r<1	1	r
	0	1-r

Each observation is presented as two observations with the same set of predictors according to the good/bad or 0/1 definition used in logistic regression. The outcome with target 1 corresponds to the rate r – weight r, the outcome with target 0 corresponds to the rate 1-r.

$$\operatorname{logit}(\mathbb{E}[Y_i \mid \mathbf{X}_i]) = \operatorname{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{\beta} \cdot \mathbf{X}_i$$

The utilization rate – one-stage, prediction period 6 month, OLS

UT and transactions behavioral predictors are significant

Limit parameter is insignificant

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	0.19868	0.00823	24.14	<.0001	0
mob	-0.00328	0.0001604	-20.44	<.0001	1.87755
limit_6	1.59E-07	1.33E-07	1.19	0.2345	2.74067
UT0_6	0.53061	0.00312	170.19	<.0001	4.59573
avg_balance_6	0.00000207	2.37E-07	8.73	<.0001	2.98617
b_AvgOB16_to_MaxOB16_ln	0.04088	0.00134	30.59	<.0001	3.08869
b_TRmax_deb16_To_Limit_ln	0.00699	0.00049122	14.22	<.0001	5.34865
b_TRavg_deb16_to_avgOB16_ln	-0.01841	0.00068774	-26.76	<.0001	3.62388
b_TRsum_deb16_to_TRsum_crd16	0.01087	0.00061894	17.55	<.0001	1.98505
b_UT1_to_AvgUT16ln	-0.00282	0.00040175	-7.01	<.0001	4.58941
b_UT1to2ln	0.00178	0.00030936	5.76	<.0001	1.91965
b_UT1to6ln	-0.0048	0.00025297	-18.99	<.0001	3.08629
b_NumDeb13to46ln	0.00545	0.00033788	16.14	<.0001	2.45803
b_inactive13	0.0824	0.00464	17.77	<.0001	5.54026
b_avgNumDeb16	0.00010706	0.00004173	2.57	0.0103	1.03309
b_OB_avg_to_eop1ln	-0.00132	0.00033009	-4	<.0001	1.15713
b_DelBucket16	0.02571	0.00235	10.96	<.0001	4.35028
b_pos_flag_0	0.01847	0.00174	10.63	<.0001	2.16061
b_pos_flag_13	0.03935	0.00213	18.51	<.0001	3.66705
b_atm_flag_0	0.053	0.00152	34.91	<.0001	1.80498
b_atm_flag_13	0.05824	0.00246	23.7	<.0001	4.5704
b_pos_flag_used46vs13	0.02783	0.0019	14.64	<.0001	1.32548
b_pos_flag_use13vs46	-0.02589	0.00203	-12.78	<.0001	1.27146
b_atm_flag_used46vs13	0.01634	0.0022	7.43	<.0001	1.88301
b_atm_flag_use13vs46	-0.02562	0.00214	-11.96	<.0001	1.34666
b_pos_use_only_flag_13	0.01185	0.00275	4.31	<.0001	2.44702
no_dpd	-0.00524	0.00413	-1.27	0.2039	3.49176
max_dpd_60	0.02693	0.00698	3.86	0.0001	1.15943
AgeGRP1	0.01494	0.00189	7.9	<.0001	2.43527
AgeGRP3	-0.00166	0.00172	-0.97	0.3323	2.3734
customer_income_ln	-0.03324	0.00165	-20.1	<.0001	2.86712
Edu_High	-0.02094	0.00175	-11.94	<.0001	2.49428
Edu_Special	-0.00198	0.00169	-1.17	0.2415	1.99299
Edu_TwoDegree	-0.01803	0.00389	-4.63	<.0001	1.24201

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
position_Man	0.01174	0.00183	6.43	<.0001	1.19365
position_Oth	0.00891	0.00181	4.91	<.0001	1.20526
position_Tech	0.00673	0.0017	3.96	<.0001	1.59924
position_Top	0.00989	0.00336	2.94	0.0033	1.10859
sec_Agricult	0.00428	0.00327	1.31	0.1905	1.09727
sec_Constr	-0.00428	0.00437	-0.98	0.3279	1.05802
sec_Energy	-0.00213	0.00281	-0.76	0.4483	1.17576
sec_Fin	-0.02624	0.00211	-12.45	<.0001	1.28055
sec_Industry	0.00344	0.00552	0.62	0.5328	1.0439
sec_Manufact	0.00805	0.0043	1.87	0.0613	1.07391
sec_Mining	0.00362	0.00299	1.21	0.2259	1.32979
sec_Service	-0.00867	0.00158	-5.49	<.0001	1.37382
sec_Trade	-0.00604	0.00212	-2.84	0.0045	1.3575
sec_Trans	-0.00676	0.00414	-1.63	0.1026	1.06506
car_Own	-0.01099	0.0015	-7.34	<.0001	1.13605
car_coOwn	0.00212	0.00228	0.93	0.3517	1.08877
real_Own	0.00051128	0.00145	0.35	0.7248	1.69724
real_coOwn	-0.00257	0.00154	-1.68	0.0939	1.5081
reg_ctr_Y	-0.00803	0.00223	-3.6	0.0003	3.21889
reg_ctr_N	-0.00653	0.00224	-2.92	0.0035	3.78825
child_1	0.01064	0.0019	5.61	<.0001	2.74174
child_2	0.0065	0.00108	6.02	<.0001	3.09289
child_3	0.03175	0.00362	8.78	<.0001	1.36959
Unempl_lnyoy_6	0.26643	0.02401	11.1	<.0001	2.16435
UAH_EURRate_lnmom_6	-0.07515	0.02953	-2.54	0.0109	1.46217
UAH_EURRate_lnyoy_6	0.16824	0.0185	9.09	<.0001	8.65389
CPI_lnqoq_6	0.62416	0.04887	12.77	<.0001	1.52323
SalaryYear_lnyoy_6	-0.26247	0.04234	-6.2	<.0001	8.33528

How customer use: POS and ATM transactions are significant

Monthly and annually changes in LCY/EUR are correlated but one is significant

Behavioural factors OLS for 3 type of models

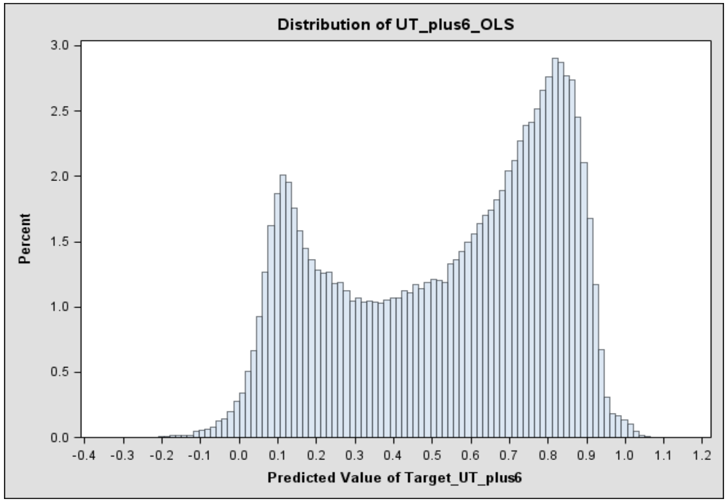
Characteristic	MOB 6+ - Limit NO Change				MOB 6+ - Limit Changed				MOB 1-5			
	Parameter Estimate	Standard error	t Value	Pr > |t|	Parameter Estimate	Standard error	t Value	Pr > |t|	Parameter Estimate	Standard error	t Value	Pr > |t|
Intercept	0.19868	0.00823	24.14	<.0001	0.14837	0.01252	11.85	<.0001	0.2773	0.0156	17.78	<.0001
Account info												
mob	-0.00328	0.0001604	-20.44	<.0001	-0.00188	0.0002648	-7.09	<.0001				
limit	1.59E-07	1.33E-07	1.19	0.2345	-2.7E-06	2.31E-07	-11.67	<.0001	-2.4E-06	2.93E-07	-8.2	<.0001
UT	0.53061	0.00312	170.19	<.0001	0.51333	0.00492	104.28	<.0001	0.43759	0.00581	75.3	<.0001
avg_balance	2.07E-06	2.37E-07	8.73	<.0001	3.84E-06	3.57E-07	10.74	<.0001	0.0000032	5.61E-07	5.7	<.0001
Behavioural - dynamic												
b_AvgOB16_to_MaxOB16_ln	0.04088	0.00134	30.59	<.0001	0.04039	0.00233	17.34	<.0001				
b_TRmax_deb16_To_Limit_ln	0.00699	0.00049122	14.22	<.0001	0.01649	0.0008275	19.93	<.0001	-0.00552	0.00045667	-12.09	<.0001
b_TRavg_deb16_to_avgOB16_ln	-0.01841	0.00068774	-26.76	<.0001	-0.03006	0.00125	-24.08	<.0001	-0.01085	0.00105	-10.31	<.0001
b_TRsum_deb16_to_TRsum_crd16_ln	0.01087	0.00061894	17.55	<.0001	0.01241	0.00119	10.46	<.0001	0.02698	0.00074873	36.04	<.0001
b_UT1_to_AvgUT16ln	-0.00282	0.00040175	-7.01	<.0001	-0.00195	0.0006926	-2.82	0.0048	0.00912	0.00068481	13.32	<.0001
b_UT1to2ln	0.00178	0.00030936	5.76	<.0001	0.0009643	0.0005288	1.82	0.0682	-0.00734	0.00021089	-34.79	<.0001
b_UT1to6ln	-0.0048	0.00025297	-18.99	<.0001	-0.00647	0.0004217	-15.34	<.0001				
b_NumDeb13to46ln	0.00545	0.00033788	16.14	<.0001	0.00937	0.0006162	15.21	<.0001				
b_inactive13	0.0824	0.00464	17.77	<.0001	0.16916	0.00812	20.84	<.0001				
b_avgNumDeb16	0.0001071	0.00004173	2.57	0.0103	-0.00149	0.0003031	-4.91	<.0001	0.00679	0.00034267	19.83	<.0001
b_OB_avg_to_eop1ln	-0.00132	0.00033009	-4	<.0001	-0.000629	0.00052	-1.21	0.2262	-0.00927	0.00064119	-14.45	<.0001
b_DelBucket16	0.02571	0.00235	10.96	<.0001	0.03251	0.00441	7.36	<.0001	0.02082	0.00821	2.54	0.0112
b_pos_flag_0	0.01847	0.00174	10.63	<.0001	0.01697	0.00244	6.94	<.0001	0.021	0.00263	7.99	<.0001
b_pos_flag_13	0.03935	0.00213	18.51	<.0001	0.04299	0.003	14.32	<.0001				
b_atm_flag_0	0.053	0.00152	34.91	<.0001	0.05427	0.00213	25.44	<.0001	0.09135	0.00279	32.7	<.0001
b_atm_flag_13	0.05824	0.00246	23.7	<.0001	0.03996	0.00355	11.25	<.0001				
b_pos_flag_used46vs13	0.02783	0.0019	14.64	<.0001	0.02612	0.0027	9.68	<.0001				
b_pos_flag_use13vs46	-0.02589	0.00203	-12.78	<.0001	-0.03041	0.00286	-10.61	<.0001				
b_atm_flag_used46vs13	0.01634	0.0022	7.43	<.0001	0.00351	0.00329	1.07	0.2861				
b_atm_flag_use13vs46	-0.02562	0.00214	-11.96	<.0001	-0.01679	0.00306	-5.48	<.0001				
b_pos_use_only_flag_13	0.01185	0.00275	4.31	<.0001	0.00375	0.00397	0.95	0.3441	-0.03498	0.0047	-7.45	<.0001
no_dpd	-0.00524	0.00413	-1.27	0.2039	0.00506	0.00694	0.73	0.4657				
max_dpd_60	0.02693	0.00698	3.86	0.0001	0.03216	0.01178	2.73	0.0063				

The utilization rate – one stage model – five methods comparison

Statistic	OLS	Fractional	Beta regression	Beta+OLS	Weighted Logistic Regrssion
Mean	0.53520	0.53893	0.50482	0.53220	0.53516
Std Deviation	0.28105	0.28297	0.25019	0.37851	0.28243
Skewness	-0.36165	-0.34751	-0.26532	-0.23848	-0.35561
Uncorrected SS	76937	78008	66833	89795	77090
Coeff Variation	52.5123	52.5070	49.5595	71.1215	52.7744
Sum Observations	112680	113465	106284	112048	112671
Variance	0.07899	0.08007	0.06259	0.14327	0.07976
Kurtosis	-1.19212	-1.37168	-1.35845	-1.59755	-1.35032
Corrected SS	16630	16859	13178	30163	16793
Std Error Mean	0.00061	0.00062	0.00055	0.00082	0.00062

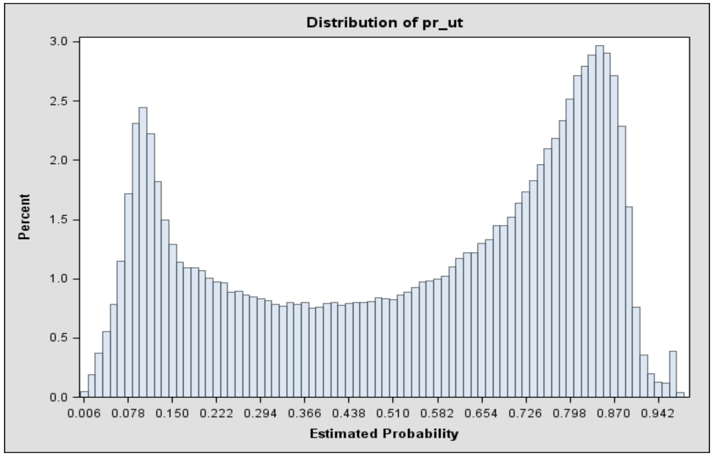
Quantile	OLS	Fractional	Beta regression	Beta+OLS	Weighted Logistic Regrssion
100% Max	1.1187	0.9821	0.9495	1.0000	0.9829
99%	0.9440	0.9251	0.8713	0.9962	0.9188
95%	0.8960	0.8845	0.8287	0.9789	0.8804
90%	0.8647	0.8626	0.8053	0.9663	0.8581
75% Q3	0.7843	0.7987	0.7346	0.9079	0.7940
50% Median	0.5985	0.6153	0.5543	0.6285	0.6103
25% Q1	0.2753	0.2551	0.2608	0.0997	0.2568
10%	0.1177	0.1152	0.1333	0.0057	0.1103
5%	0.0770	0.0904	0.1143	0.0025	0.0850
1%	-0.0045	0.0562	0.0872	0.0007	0.0446
0% Min	-0.3780	0.0055	0.0232	0.0000	0.0018

OLS distribution

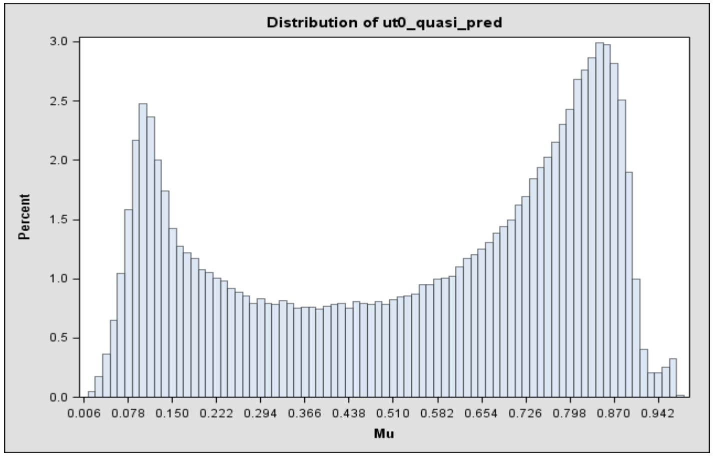


Predicted Utilization Rate Distributions for different regression methods

Weighted Logistic Regression

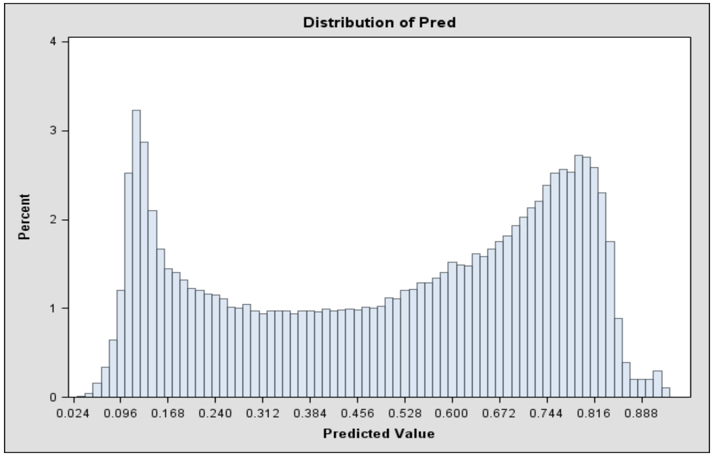


Fractional

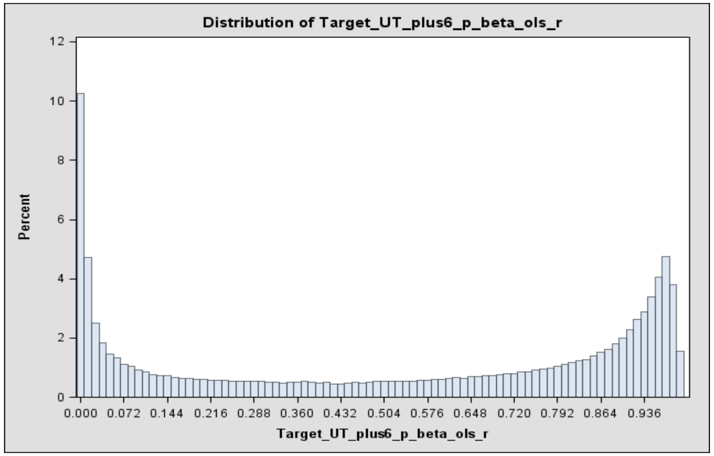


Weighted Logistic and Fractional Regression give similar distributions

Beta Regression



Beta-transformation plus OLS



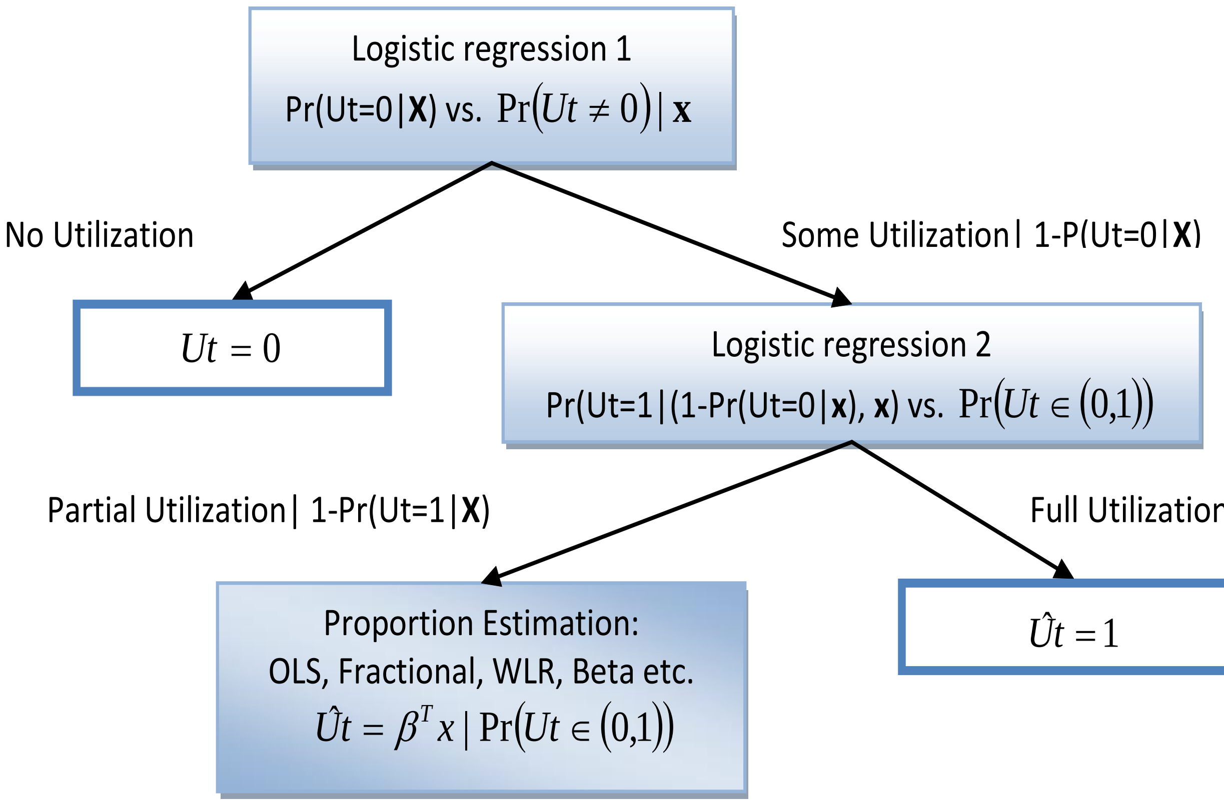
Beta-transformation has the most fitted shape, but validation results are weak

One-stage model summary

- R^2, mean absolute error (MAE), root-mean-square error (RMSE) , mean absolute percentage error (MAPE).
- More accurate results for all three segment models are given by:
  - Fractional regression
  - Weighted logistic regression
- However, OLS results are not very different.

One-Stage Model		Method	Development Sample				Validation Out-of-sample			
			R2	*MAE*	*RMSE*	*MAPE*	*R2*	*MAE*	*RMSE*	*MAPE*
Month on Book 6 or more	Limit NO Change	OLS	0.5498	0.1930	0.2537	316.3440	0.5498	0.1930	0.2537	3.1134
		Fractional (Quasi-Likelihood)	***0.5502***	0.1922	0.2544	315.9280	***0.5509***	0.1919	0.2534	313.1440
		Beta regression (nlmixed)	0.5341	0.2076	0.2589	321.1190	0.5344	0.2071	0.2580	318.4330
		Beta transformation + OLS	0.4698	0.1779	0.2761	174.0870	0.4707	0.1781	0.2751	172.0360
		Weighted Logistic Regression	**0.5522**	0.1921	0.2538	311.1860	**0.5533**	0.1917	0.2527	308.6320
	Limit Changed	OLS	0.5010	0.1967	0.2552	3.2235	0.5064	0.1955	0.2527	2.9666
		Fractional (Quasi-Likelihood)	***0.5040***	0.1950	0.2544	252.6080	***0.5099***	0.1937	0.2518	256.1710
		Beta regression (nlmixed)	0.4877	0.2080	0.2586	247.3150	0.4911	0.2071	0.2566	250.5250
		Beta transformation + OLS	0.4246	0.1831	0.2740	168.9060	0.4350	0.1810	0.2704	172.4160
		Weighted Logistic Regression	**0.5066**	0.1941	0.2538	244.2070	**0.5136**	0.1926	0.2509	247.1390
Month on Book 1-5		OLS	0.4481	0.2200	0.2820	3.2976	0.4474	0.2180	0.2802	3.1635
		Fractional (Quasi-Likelihood)	***0.4513***	0.2171	0.2812	427.6390	***0.4494***	0.2154	0.2796	421.1560
		Beta regression (nlmixed)	0.4075	0.2431	0.2922	378.2450	0.4085	0.2400	0.2898	373.4210
		Beta transformation + OLS	0.3324	0.2051	0.3102	226.6340	0.3287	0.2048	0.3088	230.1630
		Weighted Logistic Regression	**0.4535**	0.2169	0.2806	3.2565	**0.4547**	0.2146	0.2783	3.1102

Two-stage model



- Two-stage model consist of two parts:
  - the probability of zero utilization and full utilization with use of logistic regression
  - the proportion estimation with use of the set of the same methods as for one-stage model.

Two-stage model summary

At the first stage the probability that an account has zero utilization (Pr (Ut=0) and then that an account has full utilization (Pr (Ut=1)) in the performance period is calculated with binary logistic regression.

At the second stage the proportion between 0 and 1 excluding 0 and 1 values is calculated according to the set of the approaches used for one-stage direct estimation.

$$Ut = \left(1-\Pr\left(Ut=0\right)\right)\left(\Pr\left(Ut=1\right)+\left(1-\Pr\left(Ut=1\right)\right)\cdot E\left(Ut \mid Ut \neq 0, Ut \neq 1\right)\right)$$

Month on Book	Limit Changes	Stage	Method	Development Sample				Validation Out-of-sample			
Month on Book 6 and more	Limit no change	*Stage 1*	*Probability*	*KS*	*Gini*	*ROC*		*KS*	*Gini*	*ROC*	
		Pr(UT=0)	Logistic Regression	0.6262	0.7479	0.8739		0.6331	0.7547	0.8774	
		Pr(UT=1)	Logistic Regression	0.5931	0.7243	0.8622		0.6036	0.7355	0.8678	
		Stage 2	*Proportion Estimation*	*R2*	*MAE*	*RMSE*	*MAPE*	*R2*	*MAE*	*RMSE*	*MAPE*
		0<UT<1	OLS	0.4310	0.1948	0.2462	4.9151	0.4235	0.1950	0.2462	4.8260
			Fractional(Quasi-Likelihood)	0.4309	0.1946	0.2463	4.9683	0.4235	0.1950	0.2462	4.8260
			Beta regression (nlmixed)	0.4183	0.2102	0.2506	5.0499	0.4108	0.2104	0.2507	4.9075
			Beta transformation + OLS	0.3680	0.1802	0.2673	2.7377	0.3618	0.1809	0.2673	2.6513
			Weighted Logistic Regression	**0.4325**	0.1945	0.2457	4.8937	0.4253	0.1948	0.2456	4.7564
		Two-stage	*Aggregate*	*R2*	*MAE*	*RMSE*	*MAPE*	*R2*	*MAE*	*RMSE*	*MAPE*
		0<= UT <=1	OLS	0.5534	0.1913	0.2535	3.1366	0.5536	0.1910	0.2526	3.0784
			Fractional(Quasi-Likelihood)	0.5527	0.1915	0.2536	3.1590	0.5529	0.1912	0.2528	3.0979
			Beta regression (nlmixed)	0.5366	0.2068	0.2581	3.2109	0.5364	0.2063	0.2574	3.1502
			Beta transformation + OLS	0.4720	0.1773	0.2754	1.7407	0.4724	0.1774	0.2745	1.7019
			Weighted Logistic Regression	**0.5548**	0.1914	0.2531	3.1116	0.5553	0.1910	0.2521	3.0532

Conclusions

- The best validation results have been shown by for both one- and two-stage models are:
  - fractional regression
  - weighted logistic regression with data binary transformation.
- Two-stage models show slight better result for all five approaches than one-stage model: for example, $R^2 = 0.5522$ one-stage vs. 05548 two-stage for WLR
- The probabilities estimation models for the utilization rate bound values 0 and 1 have high performance results for credit risk behavioural models (GINI = 0.74 and 0.72)
- Utilization rate applied in profitability and risk models
- Other models as extensions of regressions, decision trees, survival analysis, machine learning (SVM, neural networks, etc.) can be applied and tested for further researches.

Thank you for your attention!

The Business School the University of Edinburgh

http://www.business-school.ed.ac.uk/crc

Professor Jonathan Crook

Jonathan.Crook@ed.ac.uk

Denis Osipenko, Doctoral Student

D.Osipenko@sms.ed.ac.uk

Denis.osipenko@gmail.com