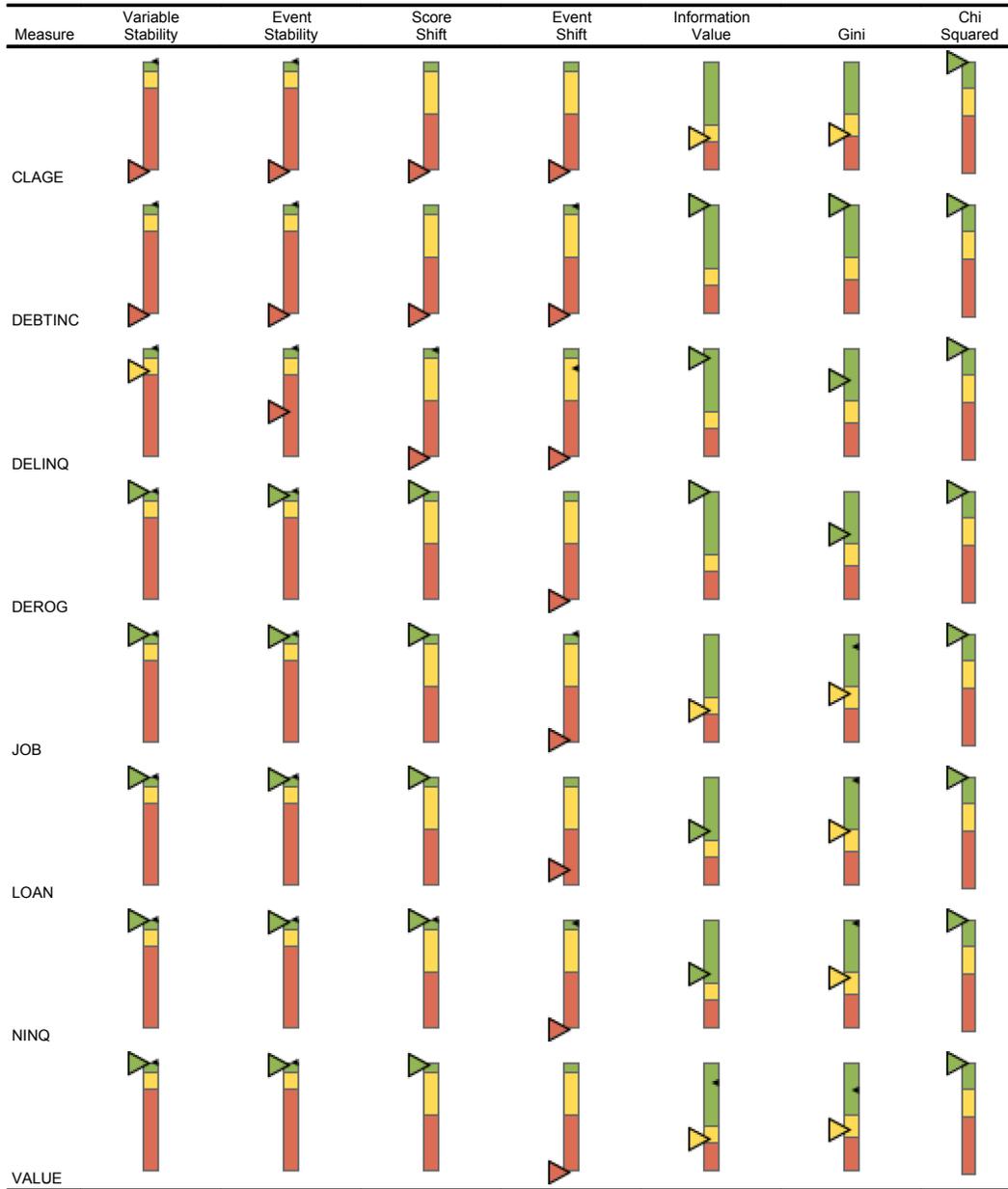


**Characteristic and attribute stability report**  
**Relative to PDA model developed on 01JAN09**  
**Mortgage product scoring performance as of 01SEP10**

**Measure stability summary**

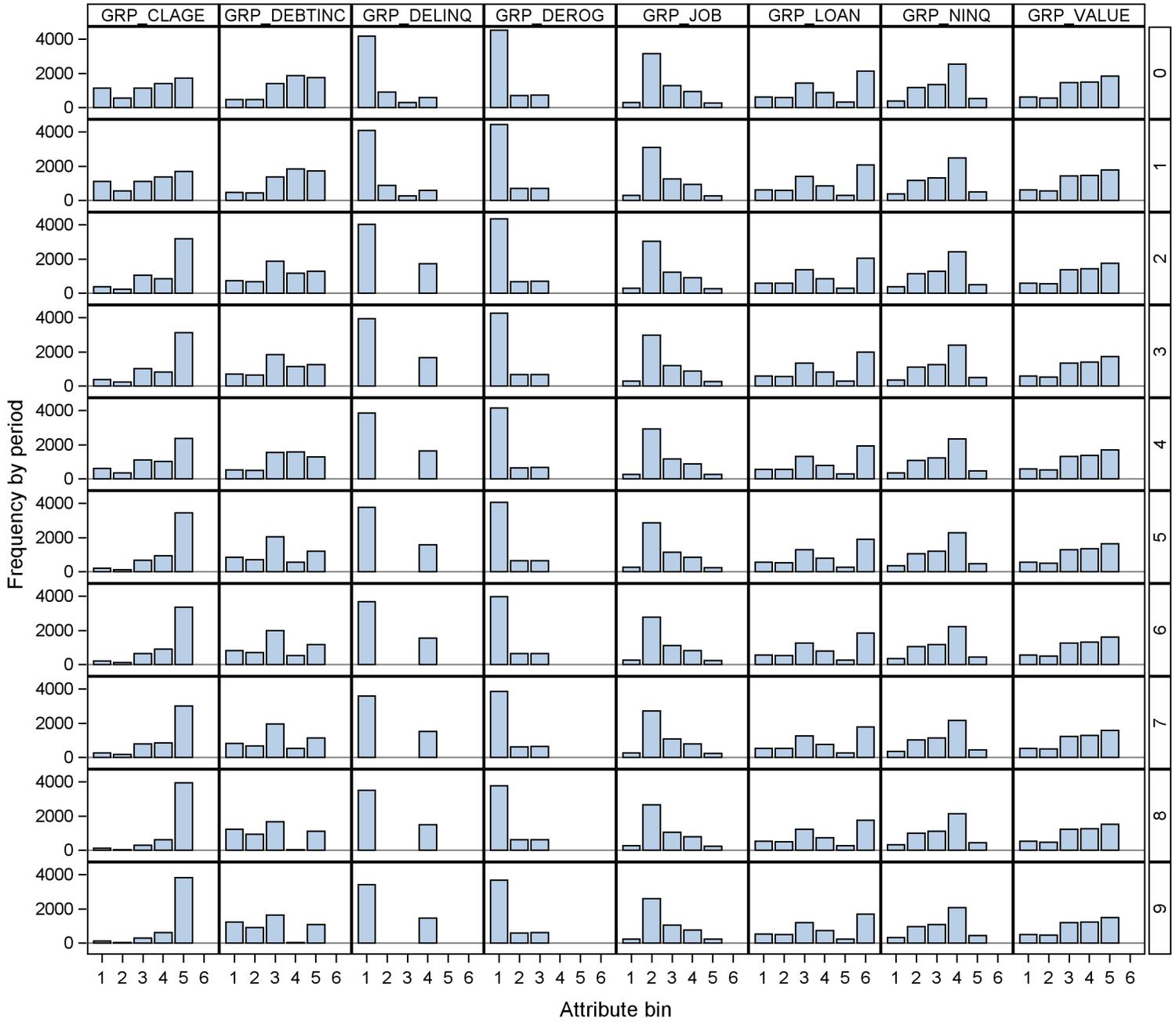


**Notes:**

The development score for Gini and Information Value is calculated based on the records for the scorecard development sample; other development score statistics are selected from the first scoring period. The actual score is calculated based on the most recent scoring sample. The KPI charts present the development score as an upwards-facing black triangle, and the actual score as a downwards-facing triangle, in the color represented by the actual score comparison with the score benchmark. The score benchmarks are tied to the statistic in question and whether the PD model is application-based or behavior-based.

**Characteristic and attribute stability report**  
**Relative to PDA model developed on 01JAN09**  
**Mortgage product scoring performance as of 01SEP10**

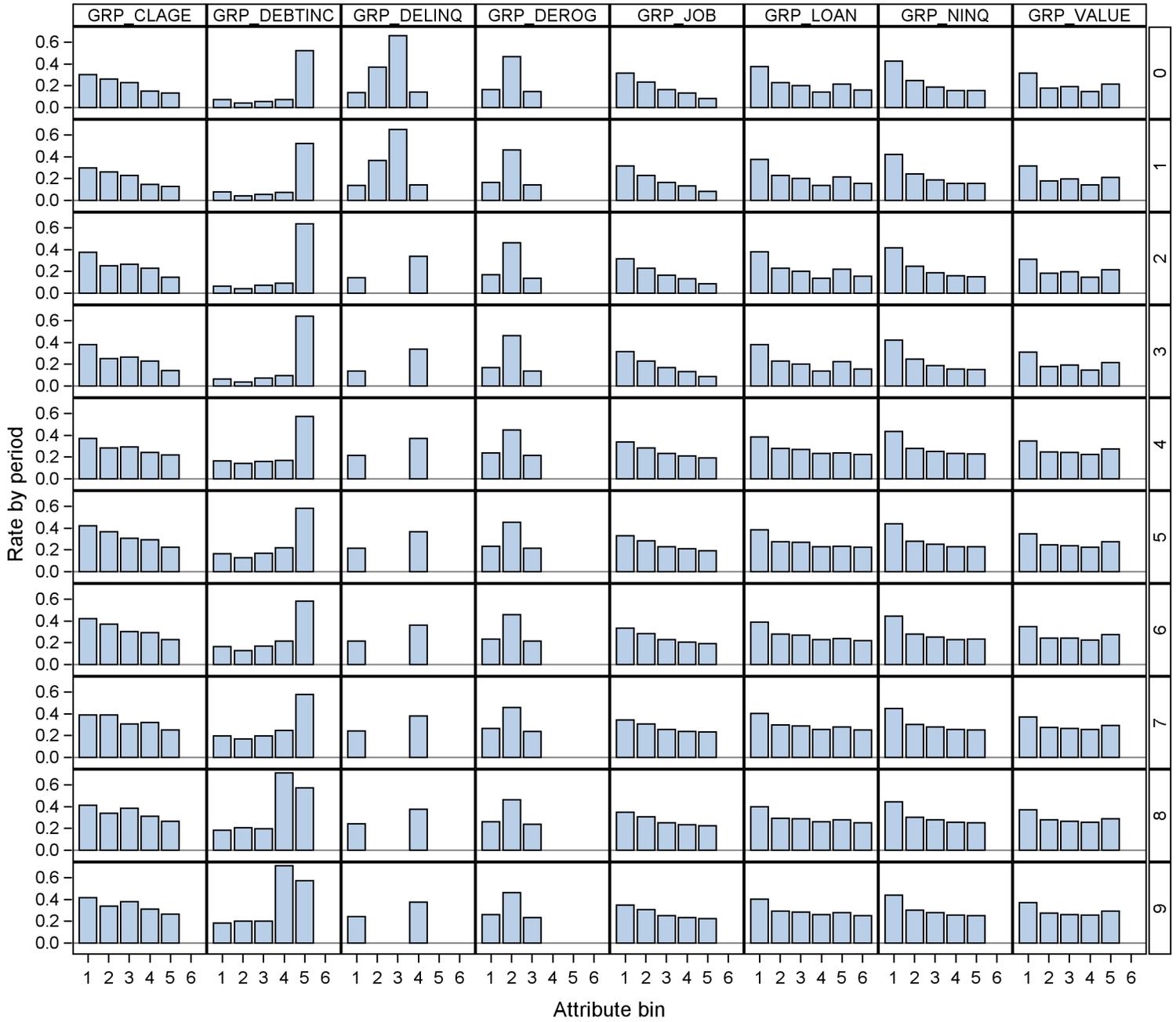
**Account distribution by characteristic and period**



Notes:  
 This analysis examines the detailed shift in characteristic and attribute count of accounts over time.

**Characteristic and attribute stability report**  
**Relative to PDA model developed on 01JAN09**  
**Mortgage product scoring performance as of 01SEP10**

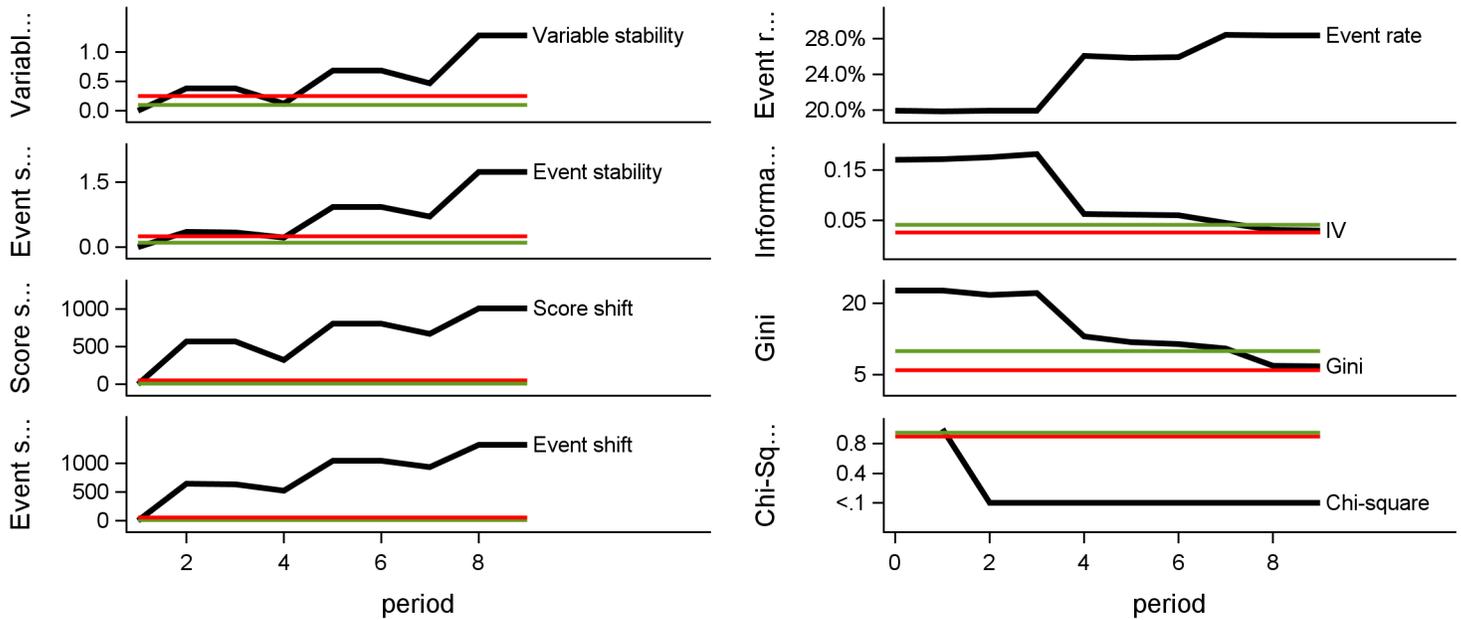
**Event rate distribution by characteristic and period**



Notes:  
 This analysis examines the detailed shift in characteristic and attribute-specific event rate over time.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Event Accounts	Event Events	Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index
CLAGE	1	1,134	341	30.1%	106	44	41.5%	0	-0.58	0.01	0.37	0.56	0.00	0.00
	2	559	146	26.1%	41	14	34.1%	2	-0.27	0.00	0.21	0.28	-17.08	-22.54
	3	1,144	264	23.1%	298	114	38.3%	14	-0.45	0.01	0.15	0.14	-183.36	-195.61
	4	1,396	211	15.1%	604	188	31.1%	25	-0.13	0.00	0.07	0.01	-276.59	-104.30
	5	1,727	227	13.1%	3,838	1,025	26.7%	30	0.08	0.01	0.49	0.74	1,486.75	1,647.47
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%		-1.36	0.03	1.29	1.74	1,009.73	1,325.01

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

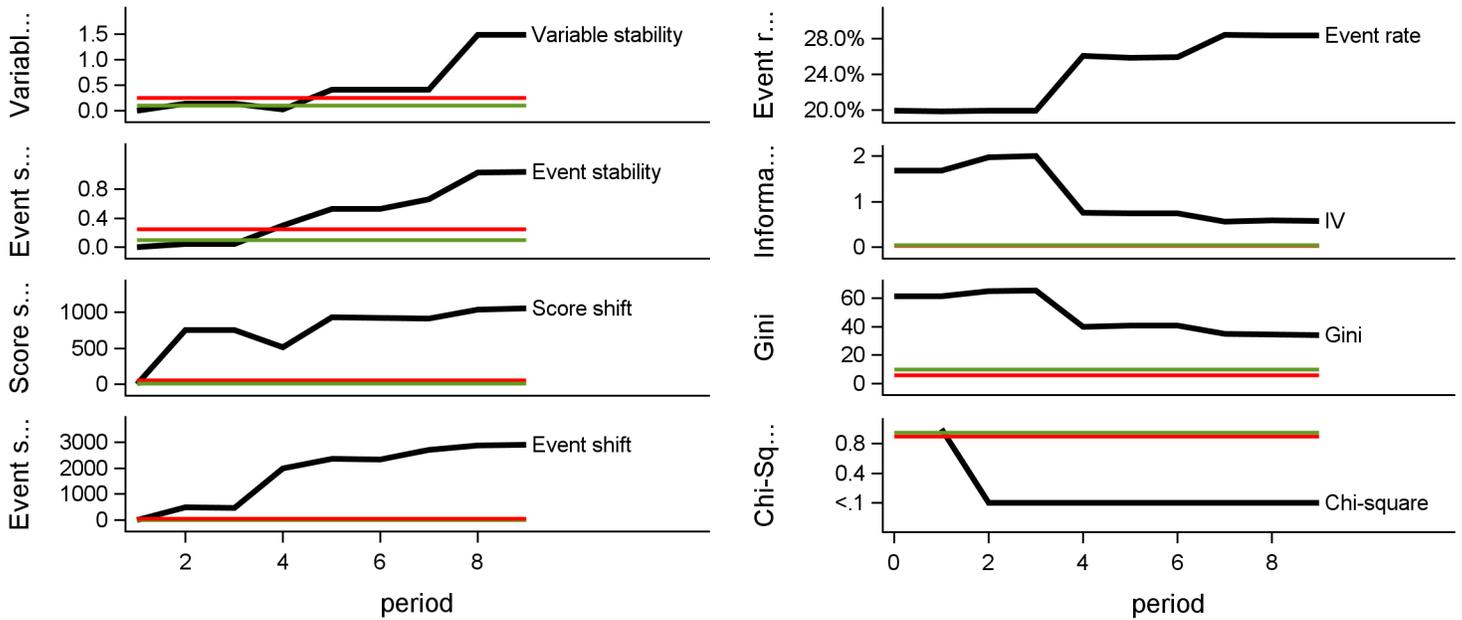
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Accounts	Events	Event rate	Score points	Weight of evidence	Informa tion value	Variable stability index	Event stability index	Score shift index	Event shift index
DEBTINC	1	467	35	7.49%	1,215	226	18.6%	48	0.55	0.07	0.20	0.23	817.26	641.95
	2	464	20	4.31%	918	186	20.3%	70	0.44	0.03	0.10	0.24	769.95	822.33
	3	1,407	78	5.54%	1,642	329	20.0%	60	0.46	0.06	0.04	0.22	599.52	1,031.66
	4	1,869	137	7.33%	38	27	71.1%	43	-1.83	0.03	1.13	0.17	-1,315.00	-411.63
	5	1,753	919	52.4%	1,074	617	57.4%	-25	-1.23	0.39	0.02	0.18	185.90	818.58
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%		-1.61	0.58	1.48	1.04	1,057.63	2,902.89

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

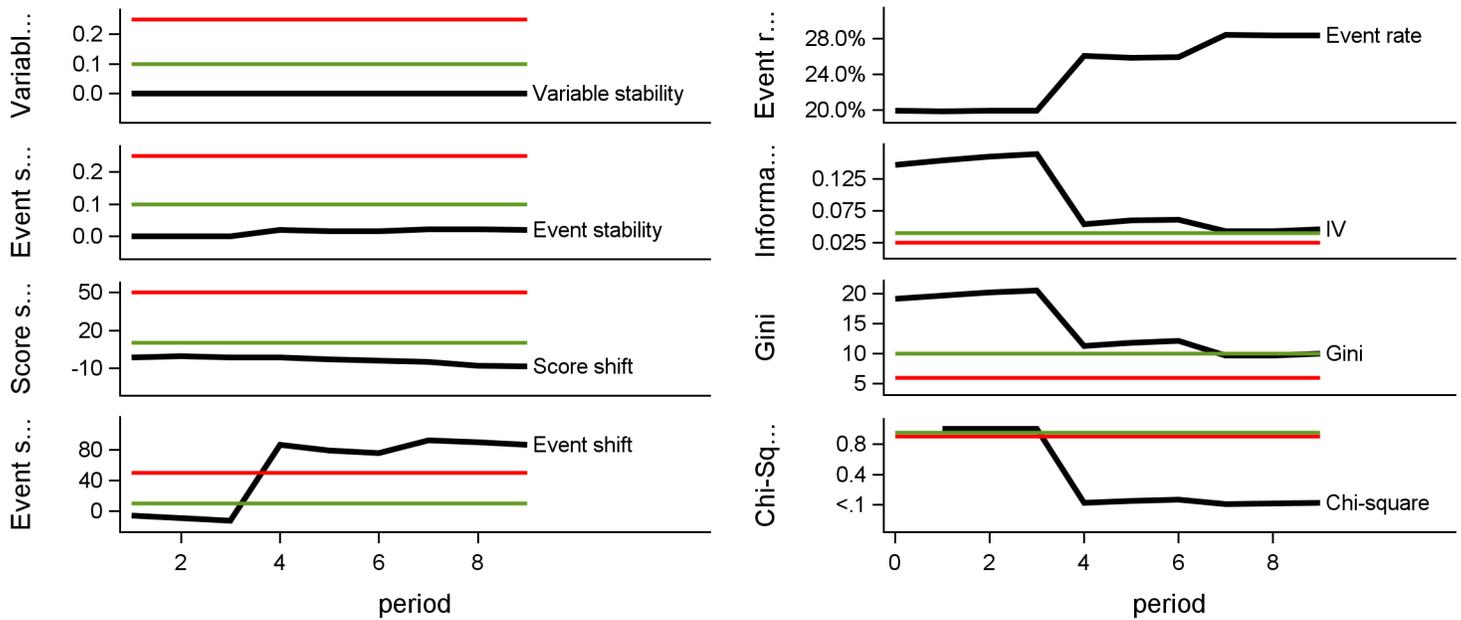
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Accounts	Events	Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index
LOAN	1	615	230	37.4%	524	212	40.5%	3	-0.54	0.03	0.00	0.01	1.21	-12.11
	2	596	136	22.8%	504	148	29.4%	12	-0.05	0.00	0.00	0.00	3.76	-9.03
	3	1,437	292	20.3%	1,188	339	28.5%	16	-0.01	0.00	0.00	0.00	3.18	-1.31
	4	870	126	14.5%	722	189	26.2%	22	0.11	0.00	0.00	0.01	3.88	67.08
	5	313	67	21.4%	249	70	28.1%	11	0.01	0.00	0.00	0.00	-1.72	-6.39
	6	2,129	338	15.9%	1,700	427	25.1%	20	0.16	0.01	0.00	0.00	-18.71	48.06
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%		-0.32	0.05	0.00	0.02	-8.40	86.30

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

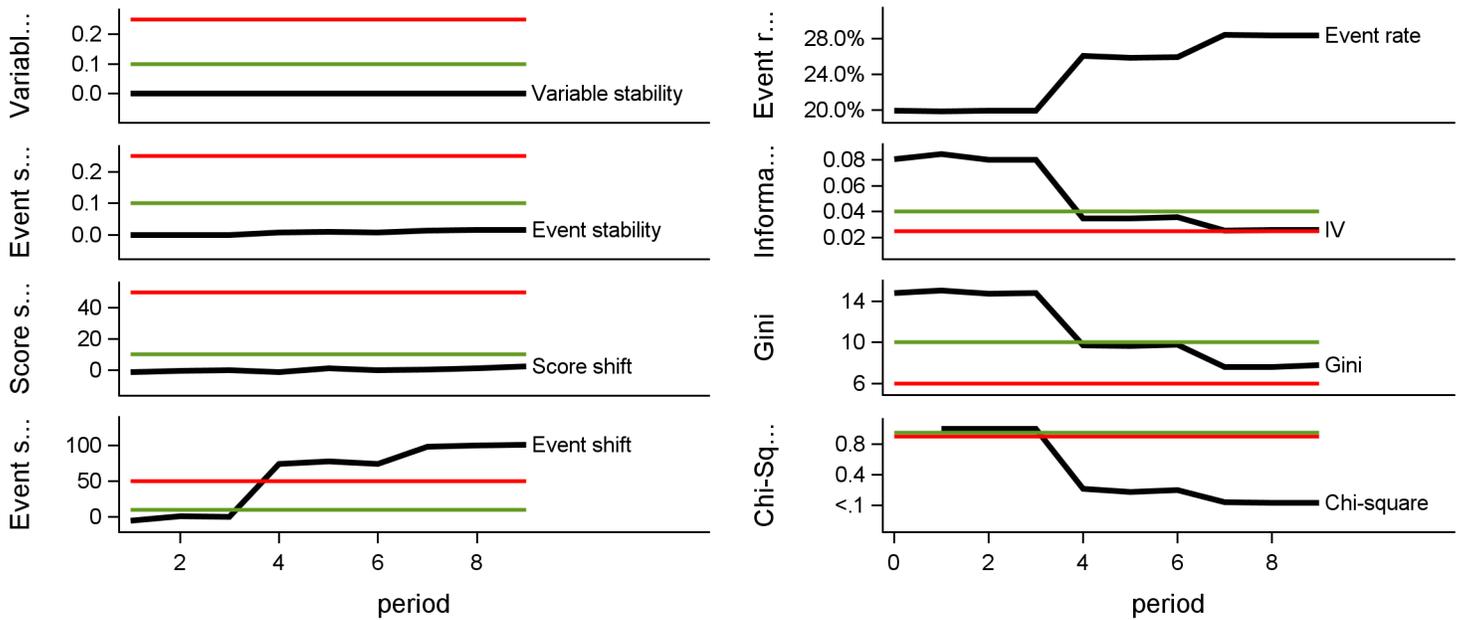
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Accounts	Events	Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index
VALUE	1	620	195	31.5%	514	191	37.2%	-1	-0.40	0.02	0.00	0.00	-0.12	2.61
	2	568	102	18.0%	469	130	27.7%	24	0.03	0.00	0.00	0.00	1.60	19.38
	3	1,453	281	19.3%	1,193	314	26.3%	17	0.10	0.00	0.00	0.00	0.55	-16.35
	4	1,490	219	14.7%	1,232	317	25.7%	26	0.13	0.00	0.00	0.01	5.45	116.20
	5	1,829	392	21.4%	1,479	433	29.3%	12	-0.05	0.00	0.00	0.00	-5.09	-20.46
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%		-0.18	0.03	0.00	0.02	2.40	101.38

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

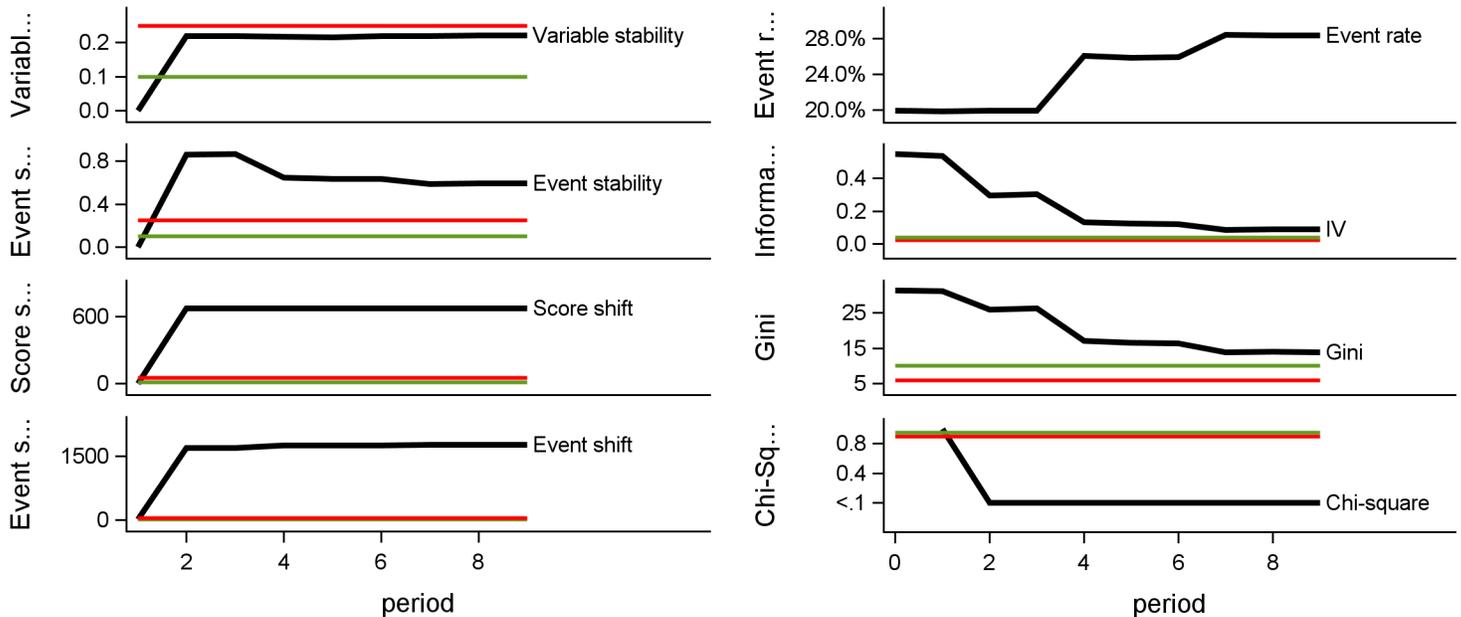
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Accounts	Events	Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index
DELINQ	1	4,179	583	14.0%	3,420	831	24.3%	28	0.21	0.03	0.00	0.02	-3.80	307.08
	4	592	84	14.2%	1,467	554	37.8%	21	-0.43	0.06	0.22	0.57	421.80	691.64
Summary		4,771	667	14.0%	4,887	1,385	28.3%		-0.22	0.09	0.22	0.59	417.99	998.72

Notes:

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

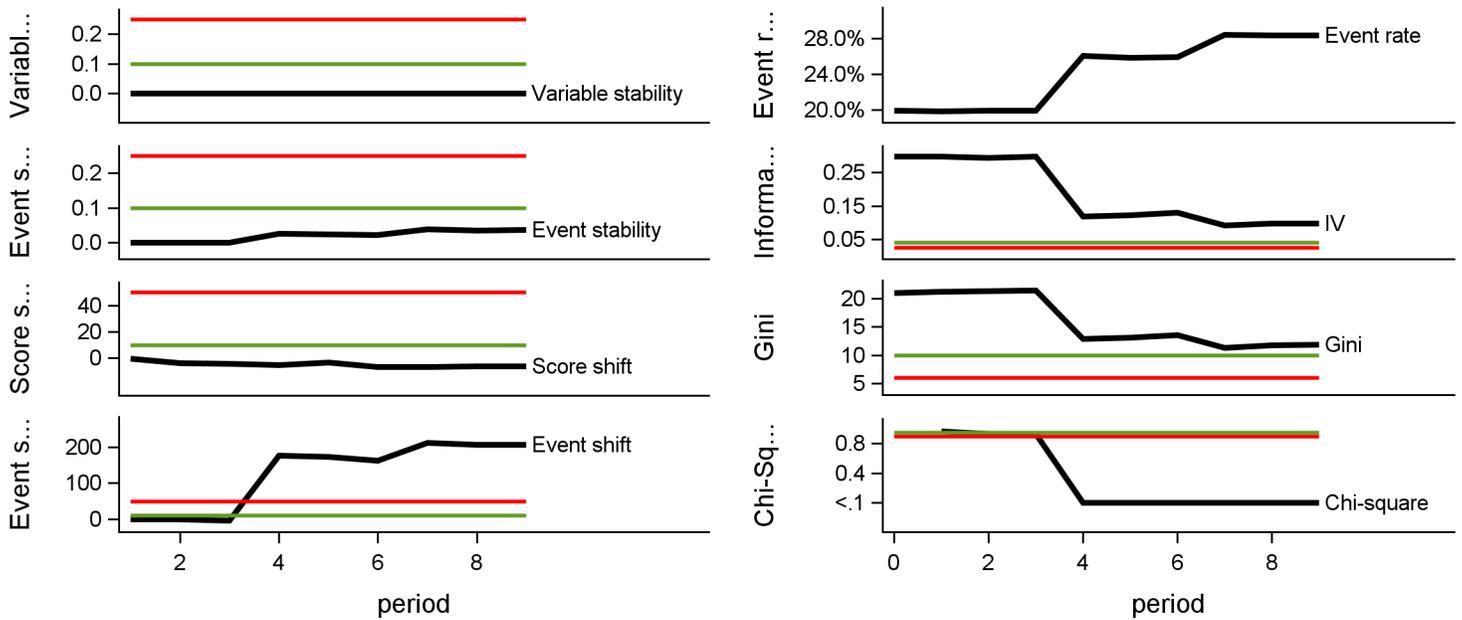
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev		Event		Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index	
		Accounts	Events	Accounts	Events									
DEROG	1	4,527	754	16.7%	3,684	969	26.3%	20	0.10	0.01	0.00	0.01	-11.45	130.99
	2	706	329	46.6%	590	272	46.1%	-6	-0.77	0.08	0.00	0.03	-1.36	48.19
	3	727	106	14.6%	613	144	23.5%	19	0.25	0.01	0.00	0.00	6.56	28.16
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%	-0.42	0.10	0.00	0.04	-6.25	207.33	

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

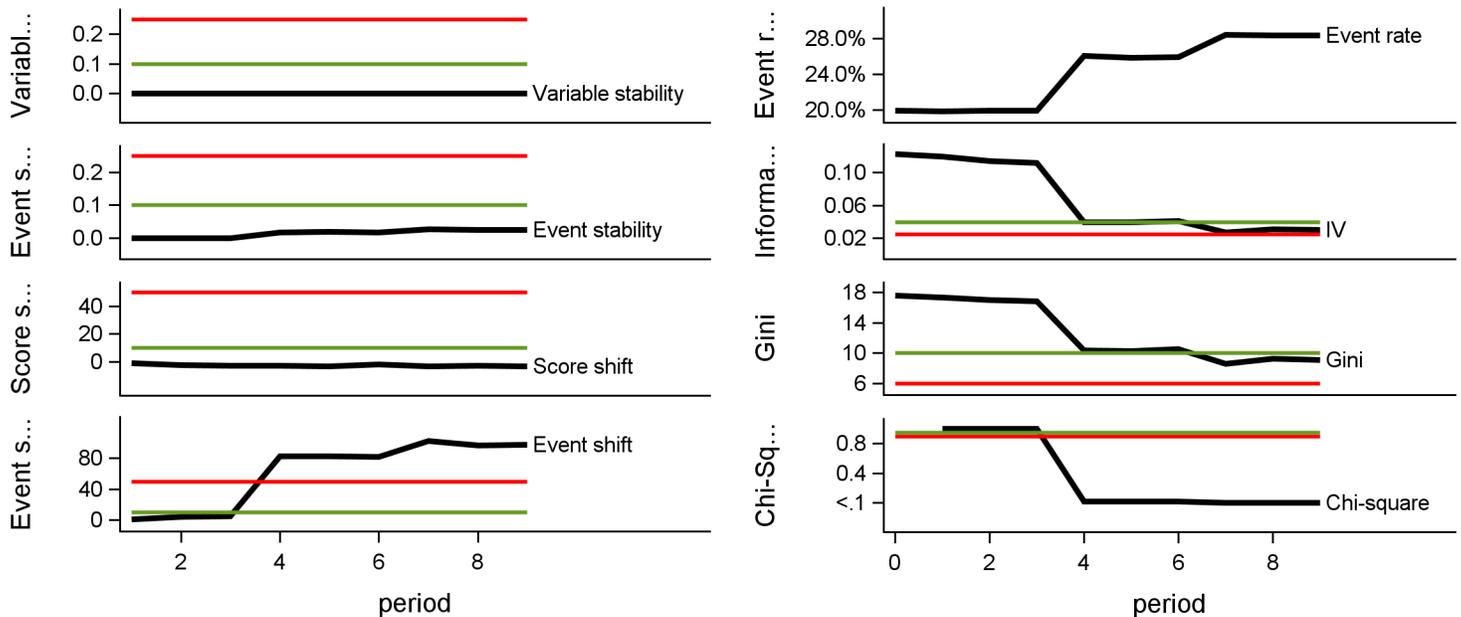
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Accounts	Events	Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index
JOB	1	302	96	31.8%	248	86	34.7%	4	-0.29	0.00	0.00	0.00	0.03	-7.46
	2	3,155	733	23.2%	2,598	801	30.8%	12	-0.12	0.01	0.00	0.00	2.70	-45.77
	3	1,276	212	16.6%	1,043	265	25.4%	20	0.15	0.00	0.00	0.00	-1.34	26.07
	4	948	125	13.2%	772	182	23.6%	25	0.25	0.01	0.00	0.01	-2.73	65.69
	5	279	23	8.24%	226	51	22.6%	34	0.31	0.00	0.00	0.01	-1.93	59.43
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%		0.29	0.03	0.00	0.03	-3.26	97.96

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

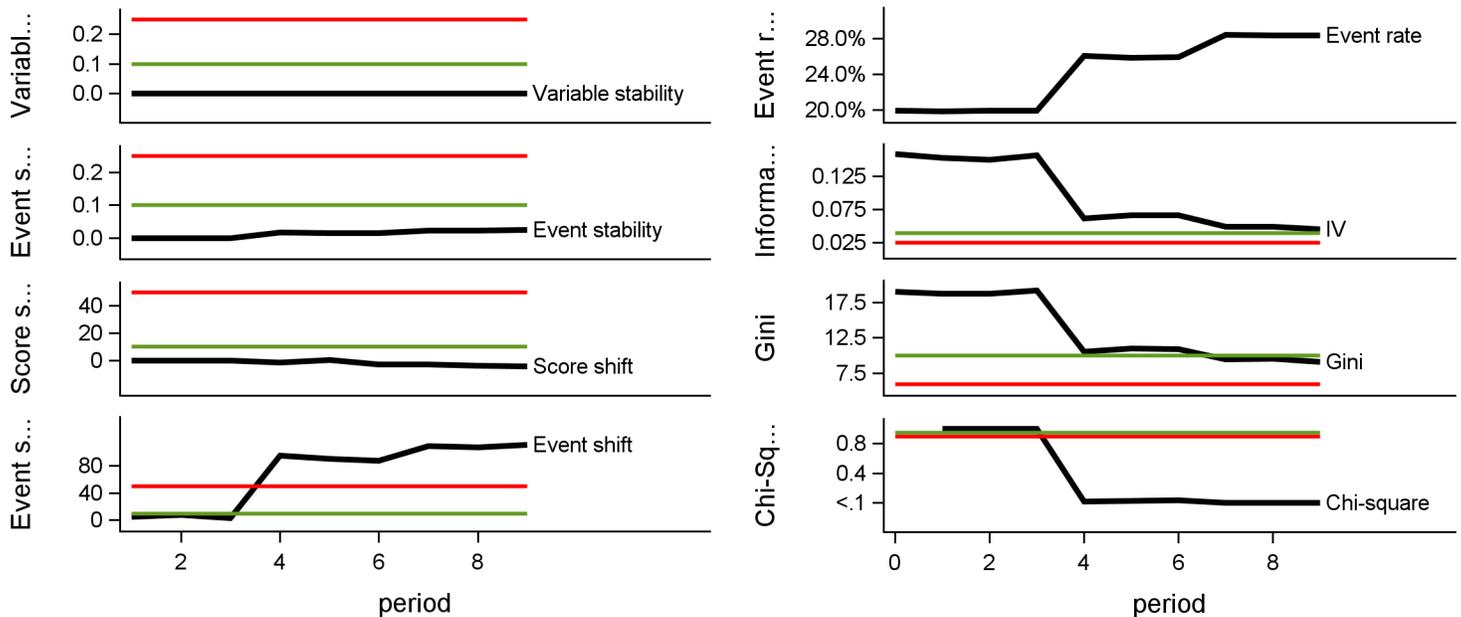
Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Characteristic and attribute stability report  
Relative to PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Attribute stability, shift and difference tests**



**Attribute metrics for development and current periods**

Characteristic	Attribute	Dev Accounts	Dev Events	Dev Event rate	Accounts	Events	Event rate	Score points	Weight of evidence	Information value	Variable stability index	Event stability index	Score shift index	Event shift index
NINQ	1	392	167	42.6%	330	145	43.9%	0	-0.68	0.04	0.00	0.01	0.00	0.00
	2	1,182	291	24.6%	977	294	30.1%	10	-0.08	0.00	0.00	0.00	1.60	-32.47
	3	1,339	254	19.0%	1,082	302	27.9%	17	0.02	0.00	0.00	0.00	-5.54	7.52
	4	2,531	396	15.6%	2,071	537	25.9%	22	0.12	0.01	0.00	0.01	-1.95	120.28
	5	516	81	15.7%	427	107	25.1%	18	0.17	0.00	0.00	0.00	1.44	16.44
Summary		5,960	1,189	19.9%	4,887	1,385	28.3%		-0.46	0.05	0.00	0.02	-4.46	111.77

**Notes:**

Event rate is ratio of events to accounts.

Weight of evidence is a measure of the relative risk of an attribute, and is defined as the log of the ratio of the nonevent rate to the event rate per attribute value.

Information value is event-to-nonevent proportion difference-weighted sum of the weight of evidence of the characteristic's attributes.

Gini index is a measure of the characteristic's ability to separate high-risk from low-risk accounts, defined by comparison of event and nonevent rates across ranked attribute bins.

Pearson chi-square quantifies the difference between the development sample and current sample based on event proportion, and the chart displays the p-value of the chi-squared test. Rejecting the test's null hypothesis means concluding that the two samples differ on event proportion.

Variable stability compares the development and current samples on distribution of the attributes of a characteristic.

Event stability compares the development and current samples on event proportion per attribute of a characteristic.

Score shift compares the development and current samples on distribution of the scorepoint-weighted attributes of a characteristic.

Event shift compares the development and current samples on event proportion per scorepoint-weighted attribute of a characteristic.

**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Model validation summary**

Test level	Test name	Development score	Actual score	Actual and Dev compared with benchmark
1: Stability	System Stability Index	<.0001	0.5237	
2: Performance	Accuracy ratio (Gini)	0.7547	0.3762	
	Kolmogorov-Smirnov statistic (K-S)	0.6205	0.3215	
	Area under the curve (AUC)	0.8773	0.6881	
	Information statistic	2.5454	0.6133	
	Kulback Leibler	1.3272	0.3505	
	(1-PH) statistic	0.9185	0.6251	
	D statistic	1.8543	0.8102	
	Pietra index	0.2194	0.1137	
	Bayesian error rate	0.1339	0.2462	
	Kendall's tau (p-value)	<.0001	<.0001	
	Somers' D (p-value)	<.0001	<.0001	
	Validation score	8.8979	3.7849	
	Accuracy	0.8601	0.7499	
	Error rate	0.1399	0.2501	
	Precision	0.6459	0.6640	
Sensitivity	0.6611	0.2383		
Specificity	0.9097	0.9523		
3: Calibration	Observed vs. estimated index	0.8691	9.2374	
	Brier score	0.0965	0.2071	
	Hosmer-Lemeshow test (p-value)	0.3747	<.0001	

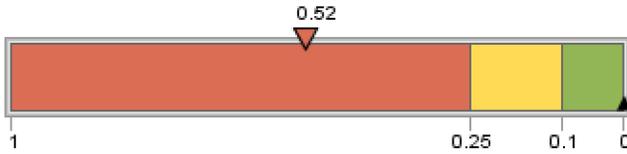
**Notes:**

The development score is calculated based on the records for the scorecard development sample. The actual score is calculated based on the most recent scoring sample. The KPI charts present the development score as an upwards-facing black triangle, and the actual score as a downwards-facing triangle, in the color represented by the actual score comparison with the score benchmark. The score benchmarks are tied to the statistic in question and whether the PD model is application-based or behavior-based.

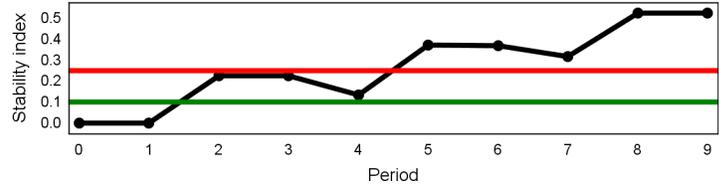
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**System stability analysis**

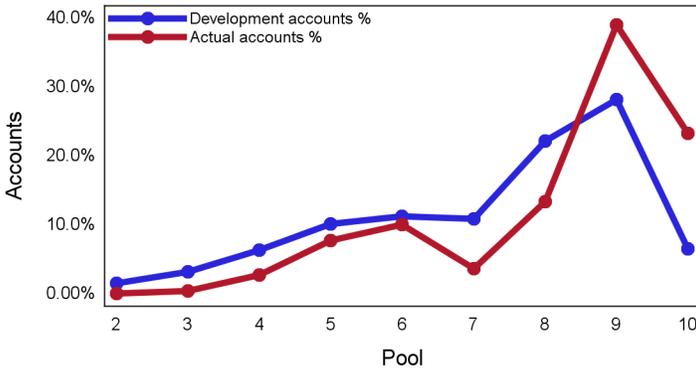
**SSI benchmarks**



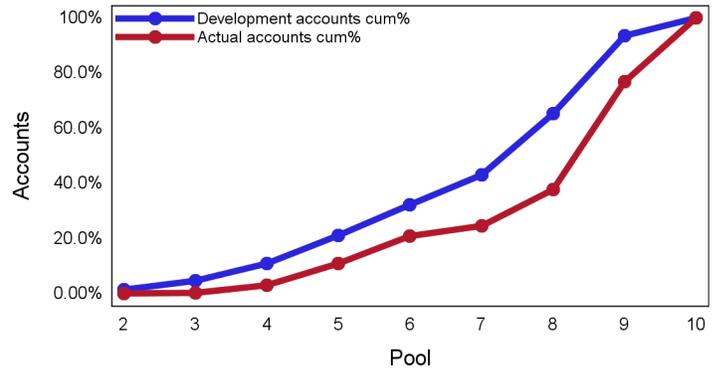
**SSI trend**



**Marginal SSI by pool**



**Cumulative SSI by pool**



**SSI metrics by pool**

Pool name	Actual accounts %	Development accounts %	Difference A-D accounts	Ratio A/D accounts	Log ratio A/D accounts	Stability index
Pool 02: 0 - 20	0.04%	1.52%	(1.48%)	2.70%	-3.61	0.05
Pool 03: 21 - 47	0.35%	3.12%	(2.77%)	11.1%	-2.19	0.06
Pool 04: 48 - 74	2.74%	6.34%	(3.60%)	43.2%	-0.84	0.03
Pool 05: 75 - 101	7.71%	10.1%	(2.39%)	76.3%	-0.27	0.01
Pool 06: 102 - 128	9.99%	11.2%	(1.23%)	89.0%	-0.12	0.00
Pool 07: 129 - 155	3.64%	10.9%	(7.22%)	33.5%	-1.09	0.08
Pool 08: 156 - 182	13.3%	22.2%	(8.82%)	60.2%	-0.51	0.04
Pool 09: 183 - 209	39.0%	28.2%	10.8%	138%	0.32	0.04
Pool 10: 210 - 238	23.2%	6.51%	16.7%	357%	1.27	0.21

**Notes:**

Economic climates and business environments change. These changes can influence population distributions and affect the performance of a scoring system. To determine if a scoring system can continue to be used effectively, the stability of the target population needs to be measured. SSI measures the degree of change within the target populations, by calculating an Index. High index values indicate that the population has changed and may suggest that the scoring system needs re-calibration or re-development. Shifts in scorecard performance by score range may demand adjustments to score cut-offs.

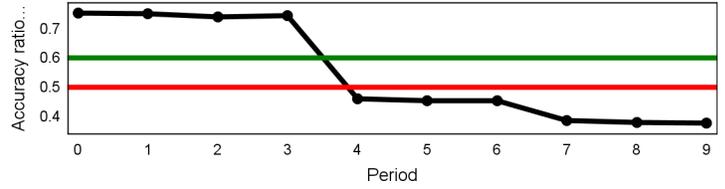
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Accuracy ratio (Gini) analysis**

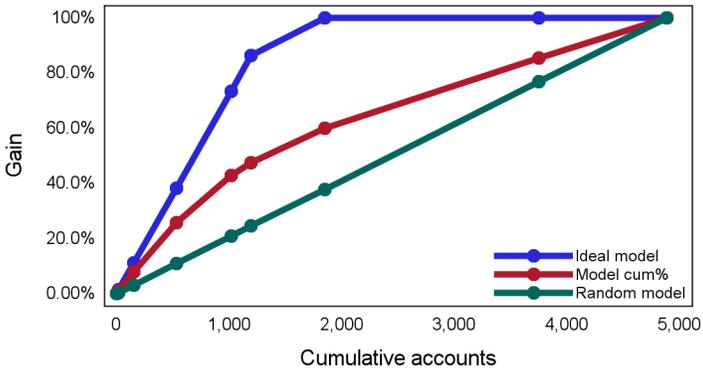
**Accuracy ratio (Gini) benchmarks**



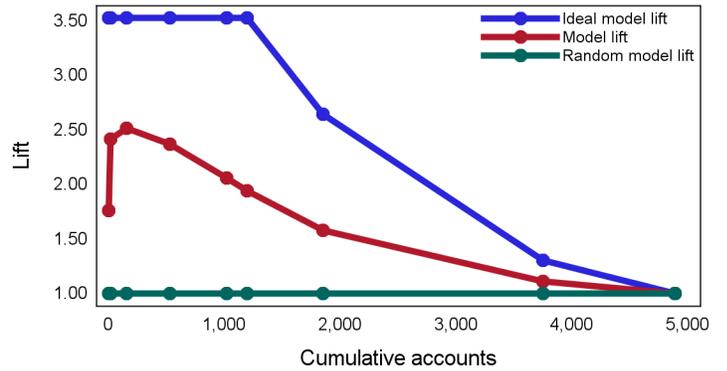
**AR (Gini) trend**



**Model gain**



**Model lift**



**Accuracy ratio (Gini) metrics by pool**

Pool name	Random model	Model cum%	Ideal model	Cumulative accounts across pools
Pool 02: 0 - 20	0.04%	0.07%	0.14%	2
Pool 03: 21 - 47	0.39%	0.94%	1.37%	19
Pool 04: 48 - 74	3.13%	7.87%	11.0%	153
Pool 05: 75 - 101	10.8%	25.7%	38.3%	530
Pool 06: 102 - 128	20.8%	42.9%	73.5%	1,018
Pool 07: 129 - 155	24.5%	47.5%	86.4%	1,196
Pool 08: 156 - 182	37.8%	59.9%	100%	1,848
Pool 09: 183 - 209	76.8%	85.4%	100%	3,752
Pool 10: 210 - 238	100%	100%	100%	4,887

**Notes:**

The gain curve is also known as the Gini curve, Power curve, or Lorenz curve. The gain curve is drawn by taking the cumulative percentage of all credit exposures on the horizontal axis and the cumulative percentage of all defaults on the vertical axis.

The Accuracy Ratio (AR) is the summary index of CAP and is also known as Gini coefficient. It shows the performance of the model being evaluated by depicting the percentage of defaults captured by the model across different scores.

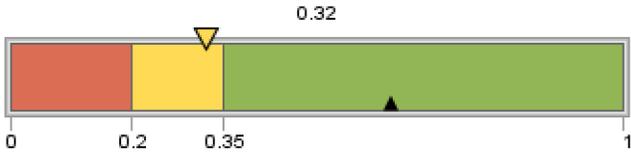
Concavity of a gain curve is equivalent to the property that the conditional probability of default, given the underlying scores, forms a decreasing function of the scores. A more concave curve indicates a better model. A perfect rating model will assign the lowest scores to the defaulters. In this case, the gain curve will increase linearly and then stay at 100%. For a random model without any discriminatory power, the percentage of all credit exposures with rating scores below a certain level (that is, the X co-ordinate) will be the same as the percentage of all defaulters with rating scores below that level (that is, the Y co-ordinate). In this case, the gain curve will be identical to the diagonal. In reality, rating systems will be somewhere in between these two extremes.

Lift charts consist of a lift curve and a baseline. The baseline reflects the effectiveness when no model is used and the lift curve reflects the effectiveness when the predictive model is used. Lift is a measure of the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model. Greater area between the lift curve and the baseline indicates a better model.

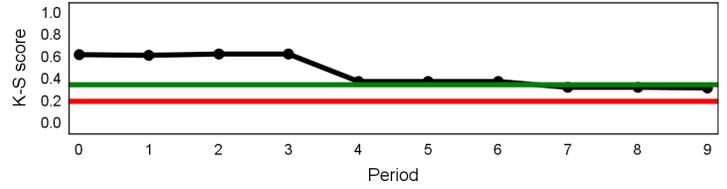
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Kolmogorov-Smirnov (K-S) score analysis**

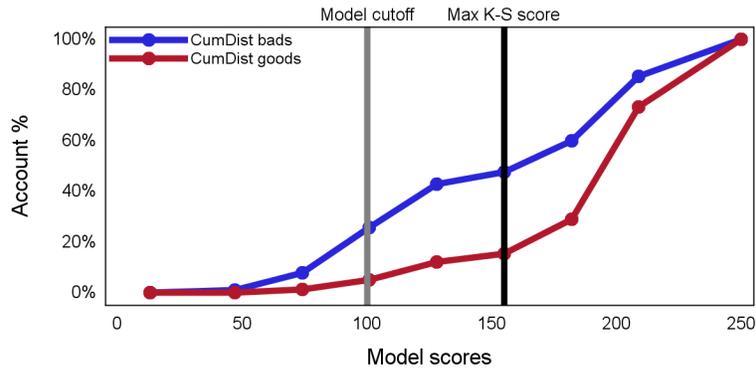
**K-S benchmarks**



**K-S trend**



**Score cutoff vs. maximum K-S**



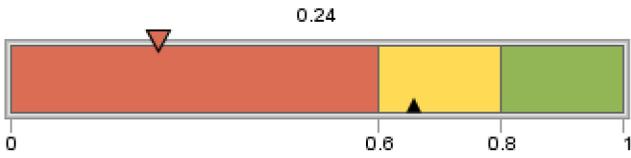
**Notes:**

The Kolmogorov-Smirnov Test is defined as the maximum distance between two population distributions. It provides the ability to distinguish healthy borrowers from troubled borrowers (usually taken as ability to discriminate defaults from non-defaults). It is also used to determine the best cut-off in application scoring. The best cut-off maximizes KS and hence it becomes the best differentiator between the two populations. The KS value can range between 0 and 100, with 100 implying the model does a perfect job in predicting defaults or separating the two populations. In general, a higher KS denotes a better model.

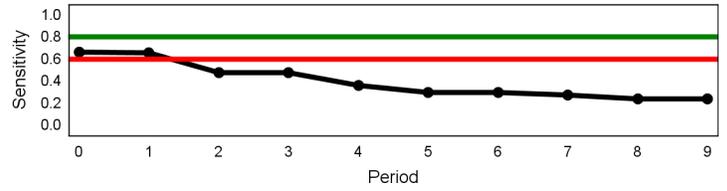
Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10

AUC and classification accuracy analysis

Sensitivity benchmarks



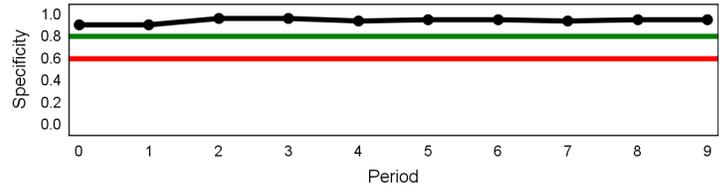
Sensitivity trend



Specificity benchmarks



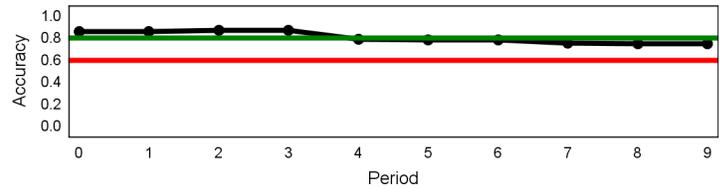
Specificity trend



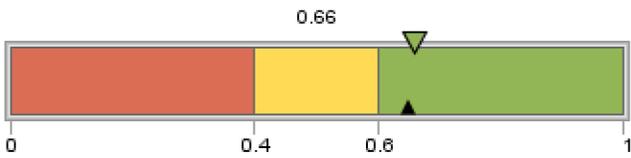
Accuracy benchmarks



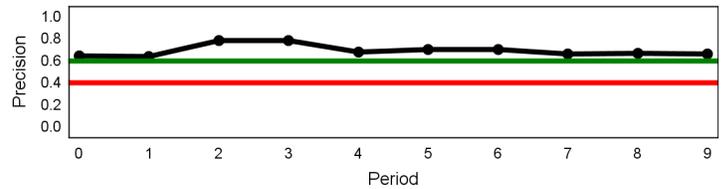
Accuracy trend



Precision benchmarks



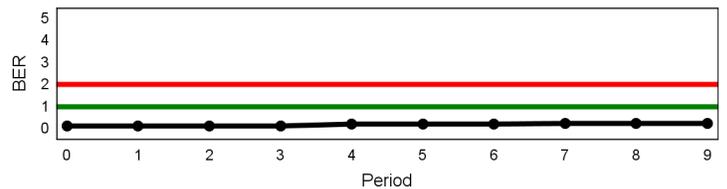
Precision trend



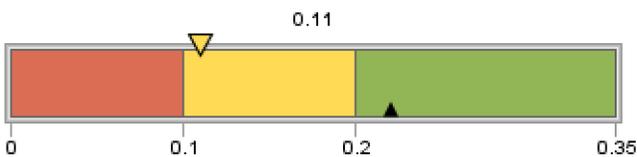
Bayesian error rate benchmarks



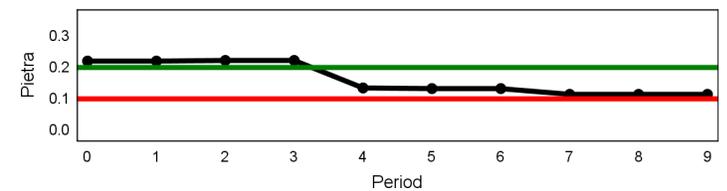
Bayesian error rate trend



Pietra index benchmarks



Pietra index trend

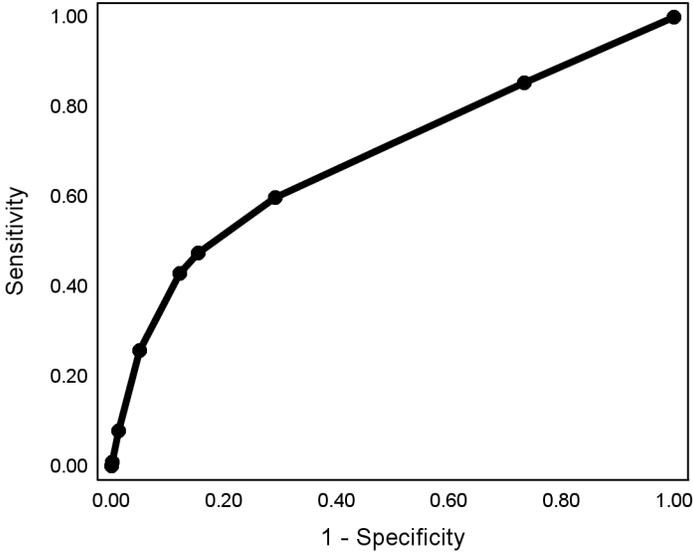


Notes: See next page for detailed description of metric definition and interpretation.

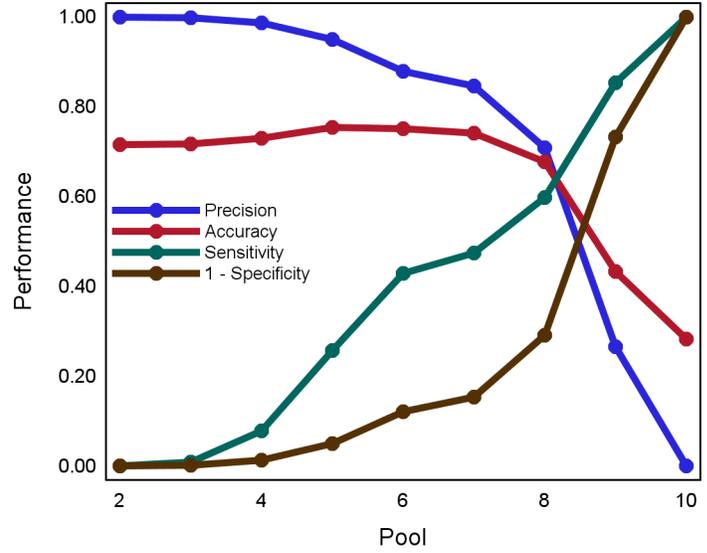
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**AUC and classification accuracy analysis**

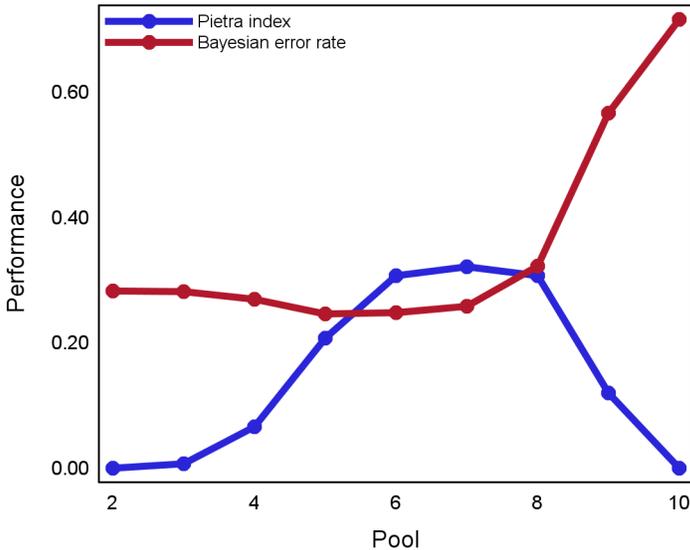
**ROC chart**



**Classification accuracy by pool**



**Pietra index and Bayesian error rate by pool**



**ROC and classification metrics by pool**

pool_name	Sensitivity	1 - Specificity	Accuracy	Precision	Bayesian error rate	Pietra index
Pool 02: 0 - 20	0.00	0.00	0.72	1.00	0.28	0.00
Pool 03: 21 - 47	0.01	0.00	0.72	1.00	0.28	0.01
Pool 04: 48 - 74	0.08	0.01	0.73	0.99	0.27	0.07
Pool 05: 75 - 101	0.26	0.05	0.75	0.95	0.25	0.21
Pool 06: 102 - 128	0.43	0.12	0.75	0.88	0.25	0.31
Pool 07: 129 - 155	0.48	0.15	0.74	0.85	0.26	0.32
Pool 08: 156 - 182	0.60	0.29	0.68	0.71	0.32	0.31
Pool 09: 183 - 209	0.85	0.73	0.43	0.27	0.57	0.12
Pool 10: 210 - 238	1.00	1.00	0.28	0.00	0.72	0.00

**Notes:**

When a bank classifies non-defaulting credits and potential defaulting credits based on the rating results for a cutoff, it is bound to incur Type I error (when a creditworthy borrower is classified as a potential defaulter (False Positive or FP)) and Type II error (when a defaulting borrower is classified as a non-defaulter (False Negative or FN)). A representation of the two errors and the two correct decisions (True Positive or TP) and (True Negative or TN) constitute the confusion matrix for that cutoff. The following statistics will be displayed based on the selected cutoff for the model.

Accuracy = (TP + TN) / (TP + TN + FP + FN); Precision = TP / (TP + FP); Sensitivity = TP / (TP + FN); Specificity = TN / (TN + FP)

Sensitivity is the ability to correctly classify defaulters that have actually defaulted. Specificity is the ability to correctly classify non-defaulters who have not defaulted. Accuracy is the proportion of the total number of predictions that were correct. Precision is the ratio of the actual defaults to the predicted defaults.

The Receiver Operating Characteristic (ROC) curve is the curve of sensitivity and '1 minus specificity' of a binary classification model for different cutoffs. A bigger area under the ROC curve (area under curve or AUC) denotes a more accurate rating model.

Area under the Curve (AUC) can be interpreted as the average ability of the rating model to accurately classify non-defaulters and defaulters. It represents the discrimination between the two populations. A higher area denotes higher discrimination. When AUC is 0.5, it means non-defaulters and defaulters are randomly classified; when AUC is 1, it means the scoring model accurately classifies non-defaulters and defaulters. Thus, in reality, the AUC ranges between 0.5 and 1.

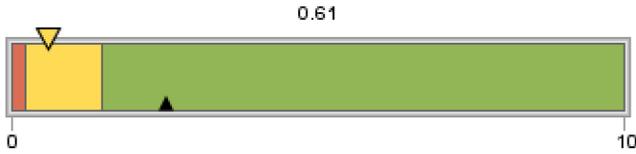
The Pietra Index can take values between zero and one. As a rating model's performance gets better, the value is closer to one. This expression is interpreted as the maximum difference between the cumulative frequency distributions of defaulters and nondefaulters.

The Bayesian error rate is the proportion of the whole sample that is misclassified when the rating system is in optimal use. For a perfect rating model, the BER will have a value of zero. A model's BER will depend on the PD. The lower the BER, lower the classification error, and better the model.

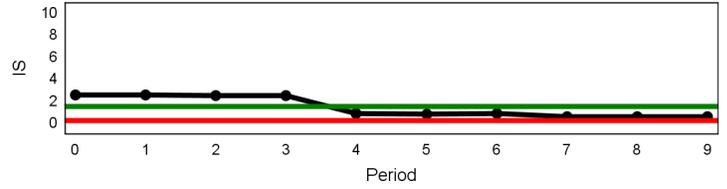
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Discrimination and association metrics analysis**

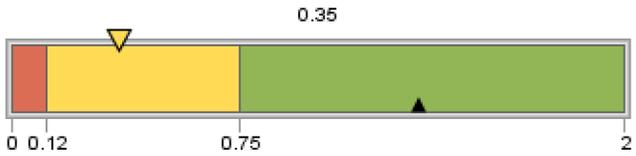
**Information statistic benchmarks**



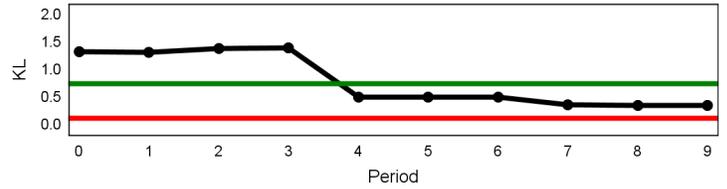
**Information statistic trend**



**Kulback Leibler benchmarks**



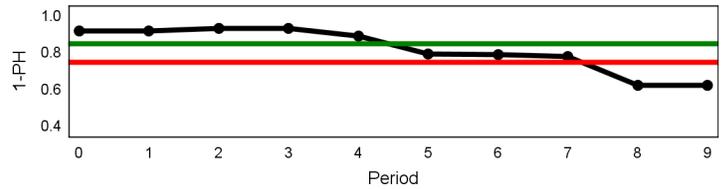
**Kulback Leibler trend**



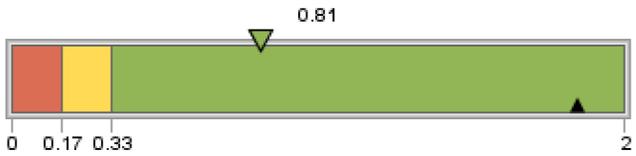
**1-PH benchmarks**



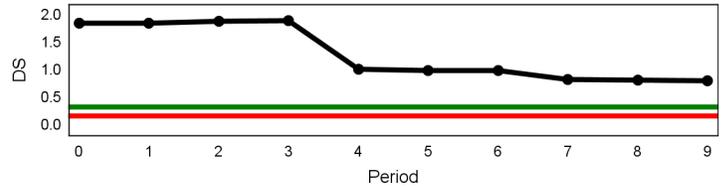
**1-PH trend**



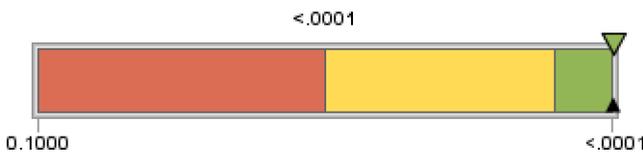
**D statistic benchmarks**



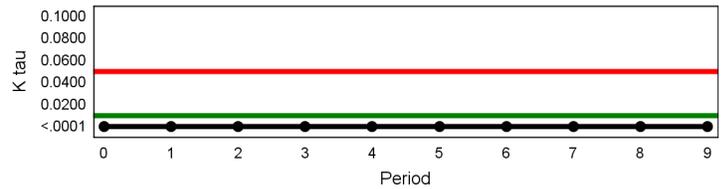
**D statistic trend**



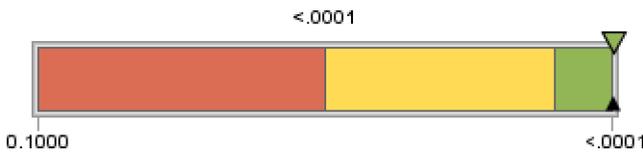
**Kendalls tau benchmarks**



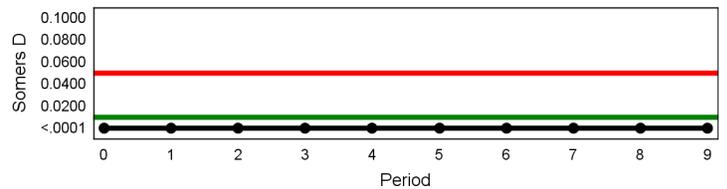
**Kendalls tau trend**



**Somers D benchmarks**



**Somers D trend**



**Notes:**

The Information Value is a weighted sum of the difference between conditional default and conditional nondefault rates. Generally speaking, the higher the information value, the more predictive a model is likely to be of default.

The Kulback-Leibler score is a non-symmetric measure of the difference between the distributions of defaulters and nondefaulters. This score has similar properties to the Information Value.

The 1-PH score is percentage of cumulative nondefaulters for the cumulative 50% of the defaulters.

The D statistic is the mean difference of scores between defaulters and nondefaulters, weighted by the relative distribution of those scores.

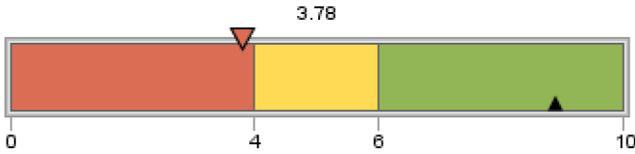
Kendall's tau-b is a nonparametric measure of association based on the number of concordances and discordances in paired observations. Kendall's tau will range between -1 and +1, with a positive correlation indicating that the ranks of both variables increase together.

Somers' D is an asymmetric modification of Kendall's tau. Somers' D differs from tau-b in that it uses a correction only for pairs that are tied on the independent variable. It will also range between -1 and +1.

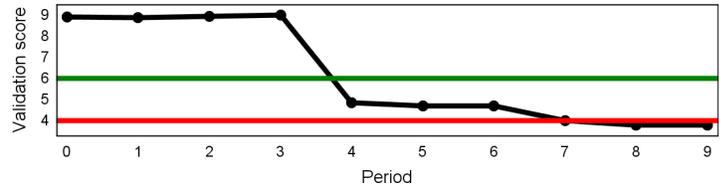
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Validation score analysis**

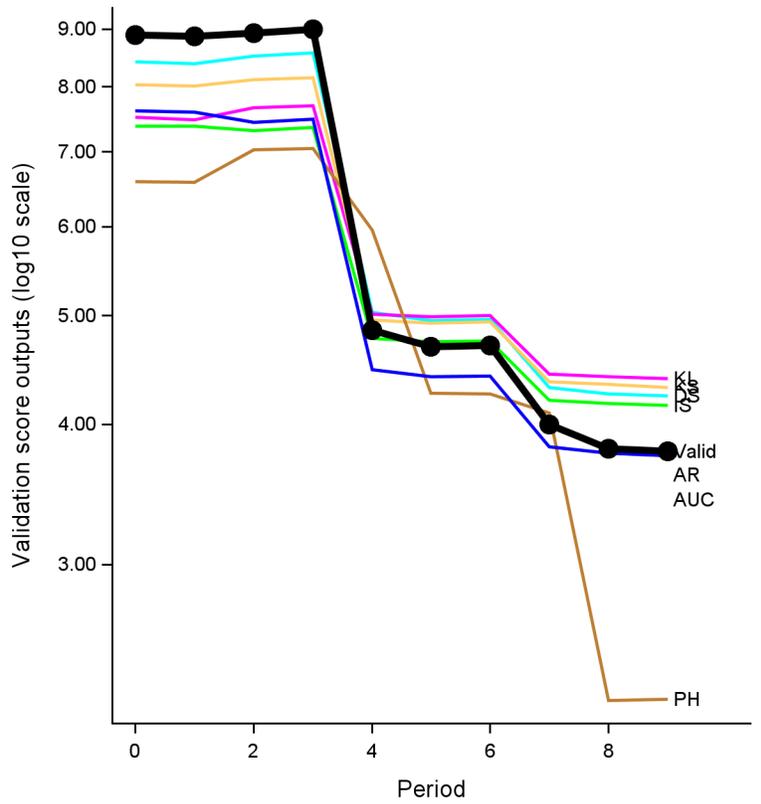
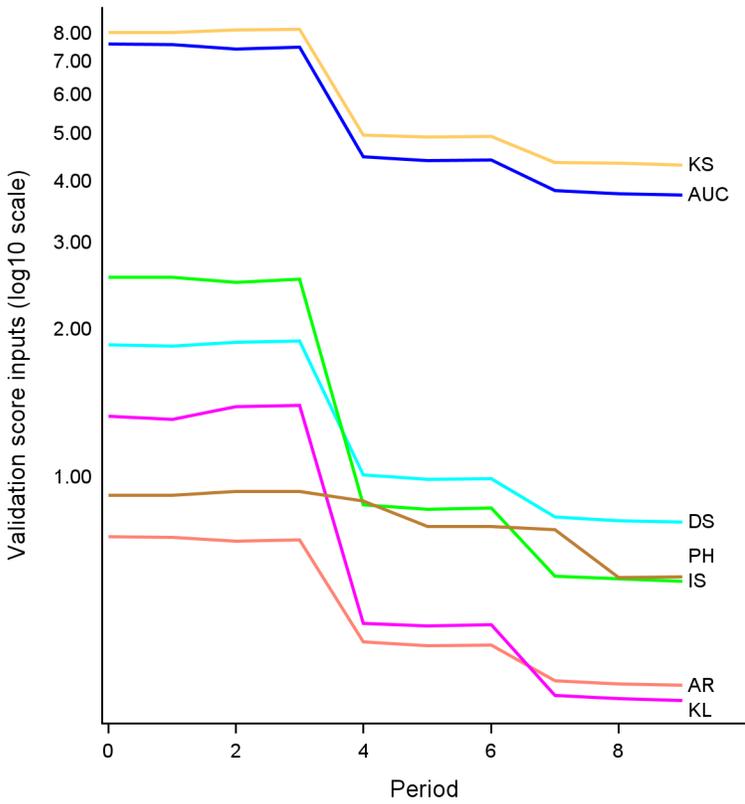
**Validation score benchmarks**



**Validation score trend**



**Comparison of validation score inputs and outputs**



**Notes:**

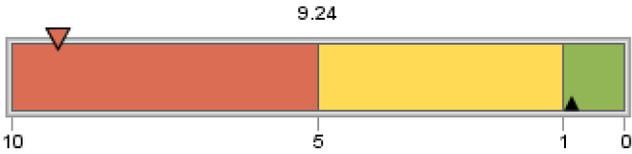
Validation score is the average scaled value of seven distance measures, anchored to a scale of 1 to 13, lowest to highest. The seven measures used include mean difference (D), percentage of cumulative nondefaulters for the cumulative 50% of the defaulters (PH), Maximum deviation (KS), Gini Coefficient (G), Information statistic (I), area under the ROC statistic (AUC) and Kullback-Leibler statistic (KL).

The rationale behind this approach is that different distance measures may rank scoring systems in different sequences, so we can use a blended validation score for decisions instead of relying on just one measure that could lead to conflicting results.

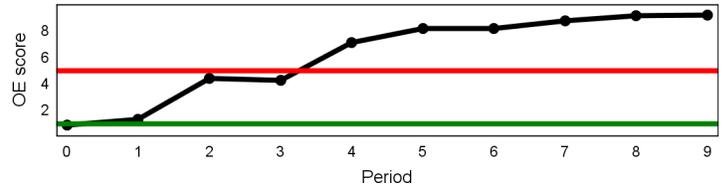
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Observed vs. estimated PD analysis**

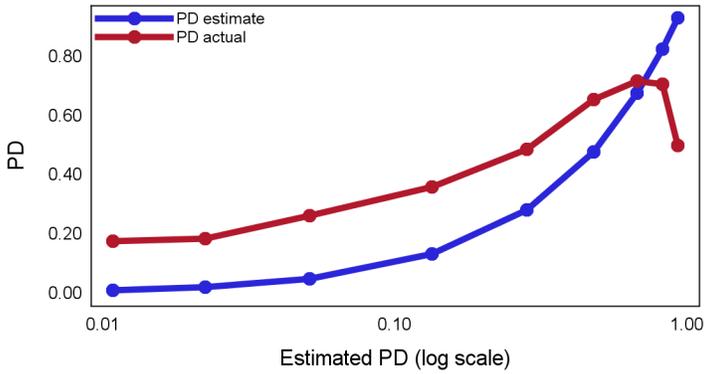
**Observed vs. estimated index benchmarks**



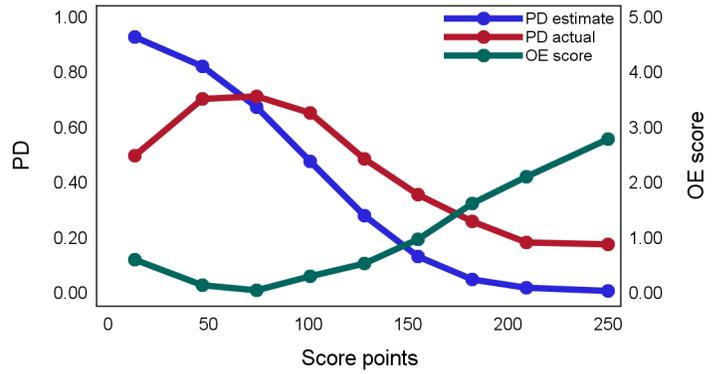
**Observed vs. estimated PD trend**



**Estimated vs. observed PD**



**Observed PD and OE index by score points**



**Observed vs estimated PD by pool**

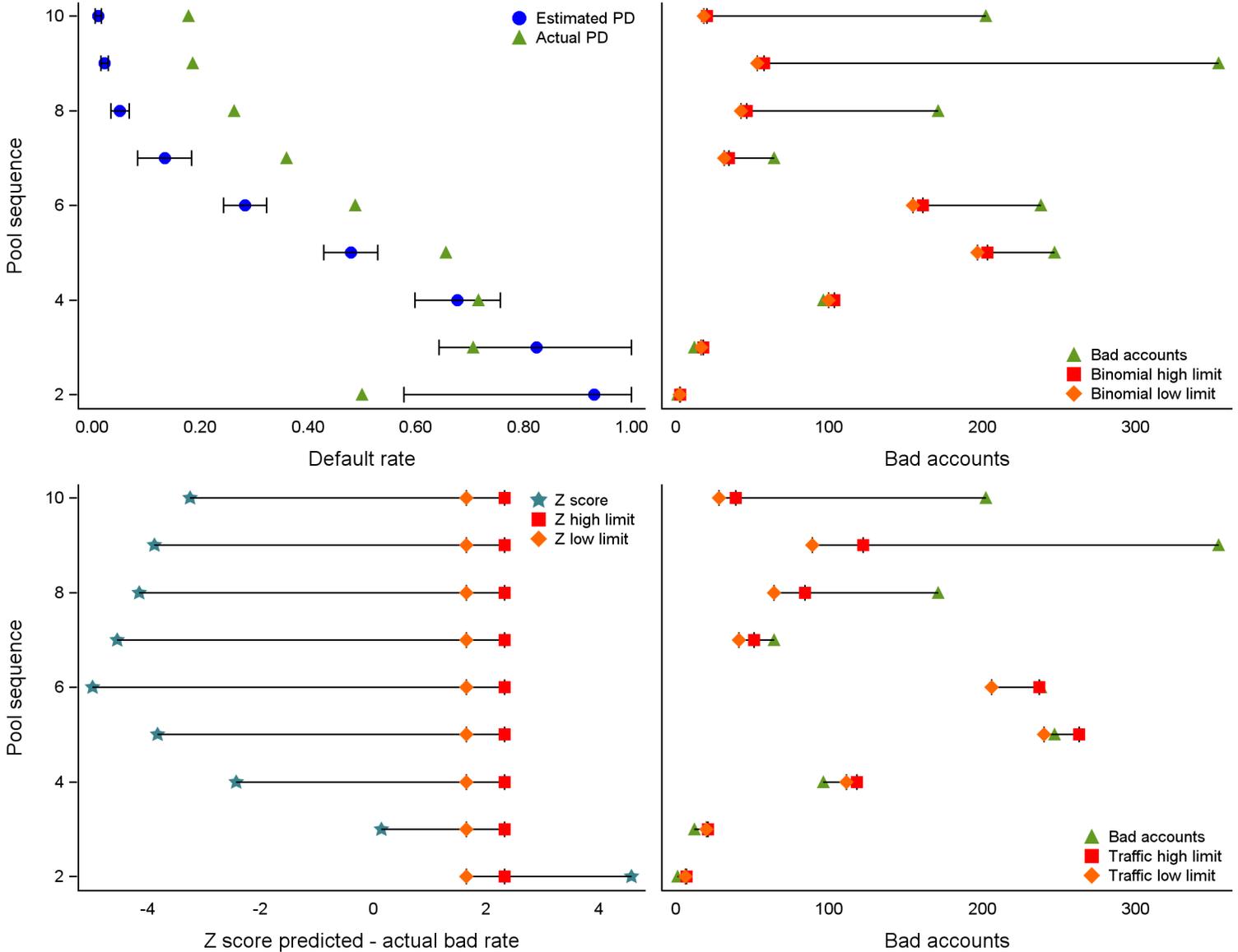
Pool	Score points	PD estimate	PD actual	OE score
Pool 02: 0 - 20	13	0.93	0.50	0.62
Pool 03: 21 - 47	47	0.82	0.71	0.15
Pool 04: 48 - 74	74	0.68	0.72	0.06
Pool 05: 75 - 101	101	0.48	0.66	0.31
Pool 06: 102 - 128	128	0.28	0.49	0.54
Pool 07: 129 - 155	155	0.13	0.36	0.99
Pool 08: 156 - 182	182	0.05	0.26	1.64
Pool 09: 183 - 209	209	0.02	0.19	2.12
Pool 10: 210 - 238	250	0.01	0.18	2.80

**Notes:**

The observed vs. estimated index is a measure of closeness of the observed and estimated default rates. Hence, it measures the model's ability to predict default rates. The closer the index is towards zero (0), the better the model performs in predicting default rates.

**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Confidence interval analysis and binomial/normal/traffic light tests**



**Notes:**

The four charts above plot measures of actual default rates with various expectations given predictive model properties. The upper-left chart compares actual PD rates per pool with estimated PD rates and the confidence limits of the estimate. If the estimated PD lies in the CI limits of the actual PD model, it performs better in estimating actual outcomes.

The upper-right chart compares the number of bad accounts per pool with the two limits estimated via the binomial test. This test evaluates whether the PD of a pool is underestimated. If the number of bad accounts per pool exceeds either the low limit (binomial test at 90% confidence) or high limit (binomial test at 95% confidence) it suggests the model is poorly calibrated.

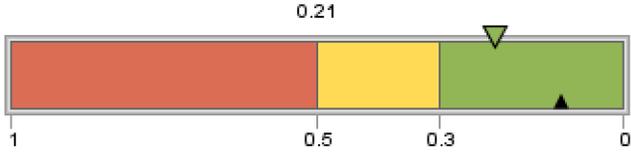
The lower-left chart compares the normalized difference of predicted and actual bad rates per pool with two limits estimated over multiple observation periods. This test measures the pool stability across time. If a majority of the pools lie in the rejection region (to the right of the limits) then the pooling strategy should be revisited.

The lower-right chart compares the number of bad accounts per pool with the two limits estimated via the traffic light test. This test evaluates whether the PD of a pool is underestimated, but unlike the binomial test, it does not assume that cross-pool performance is statistically independent. If the number of bad accounts per pool exceeds either the low limit (traffic test at 90% confidence) or high limit (traffic light test at 95% confidence) it suggests the model is poorly calibrated.

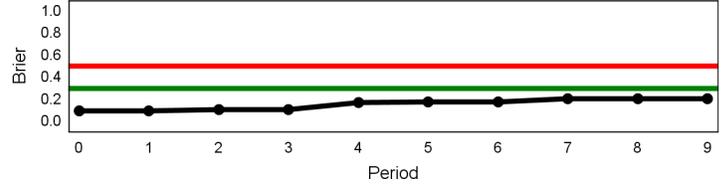
**Model validation report: Application-level probability of default  
PDA model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Calibration metrics analysis**

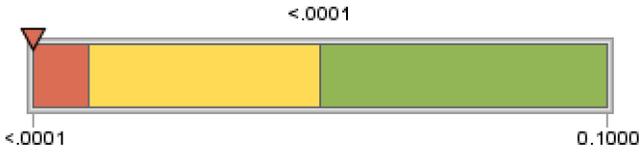
**Brier score benchmarks**



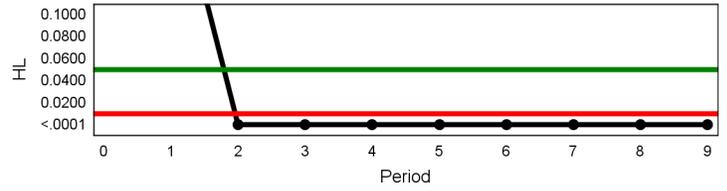
**Brier score trend**



**Hosmer Lemeshow (p-value) benchmarks**



**Hosmer Lemeshow (p-value) trend**



**Notes:**

The Brier score measures the accuracy of probability assessments at the account level. It measures the average squared deviation between predicted probabilities for a set of events and their outcomes, so a lower score represents higher accuracy.

The Hosmer–Lemeshow test is a statistical test for goodness of fit for classification models. The test assesses whether or not the observed event rates match expected event rates in pools. Models for which expected and observed event rates in pools are similar are well calibrated. The p-value of this test is a measure of the accuracy of the forecasts of default probabilities; the closer the p-value is to zero, the more poorly calibrated is the model.

**Model validation report: Application-level loss given default  
LGD model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Model validation summary**

Test level	Test name	Development score	Actual score	Actual and Dev compared with benchmark
0: Stability	System Stability Index	<.0001	<.0001	
1: Performance	Mean Squared Error (MSE)	0.0094	0.0134	
	Mean Absolute Deviation (MAD)	0.0595	0.0761	
	Mean Absolute Percent Error (MAPE)	0.0013	0.0034	
	Correlation Analysis	0.8585	0.8156	
2: Calibration	Accuracy Ratio (Gini)	0.2886	0.2326	

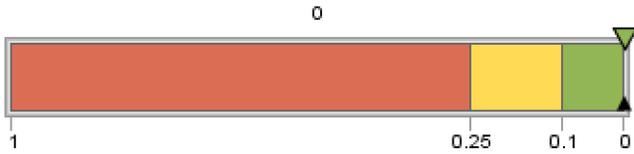
**Notes:**

The development score is calculated based on the records for the scorecard development sample. The actual score is calculated based on the most recent scoring sample. The KPI charts present the development score as an upwards-facing black triangle, and the actual score as a downwards-facing triangle, in the color represented by the actual score comparison with the score benchmark. The score benchmarks are tied to the statistic in question and whether the LGD model is application-based or behavior-based.

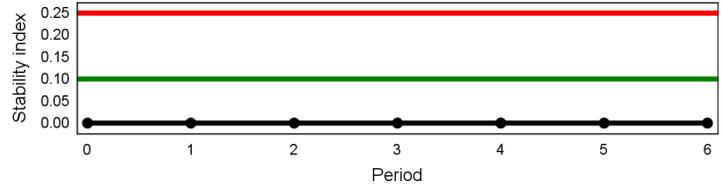
**Model validation report: Application-level loss given default  
LGD model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**System stability analysis**

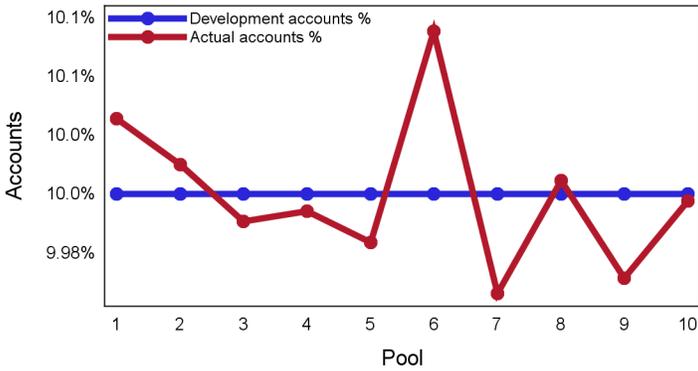
**SSI benchmarks**



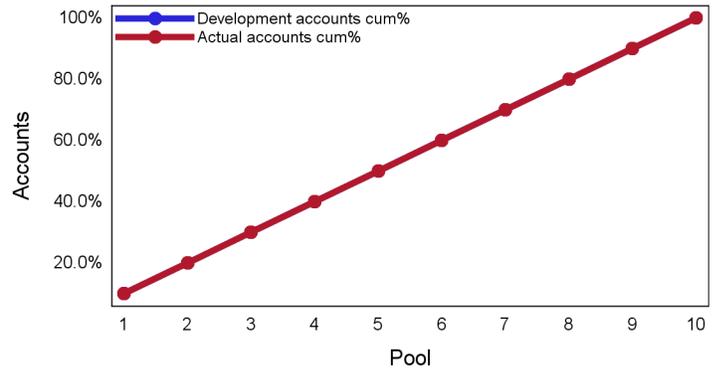
**SSI trend**



**Marginal SSI by pool**



**Cumulative SSI by pool**



**SSI metrics by pool**

Pool name	Actual accounts %	Development accounts %	Difference A-D accounts	Ratio A/D accounts	Log ratio A/D accounts	Stability index
Pool 01: 0.45 - 1.00	10.0%	10.0%	0.03%	100%	0.00	0.00
Pool 02: 0.28 - 0.45	10.0%	10.0%	0.01%	100%	0.00	0.00
Pool 03: 0.19 - 0.28	9.99%	10.0%	(0.01%)	99.9%	-0.00	0.00
Pool 04: 0.16 - 0.19	9.99%	10.0%	(0.01%)	99.9%	-0.00	0.00
Pool 05: 0.13 - 0.16	9.98%	10.0%	(0.02%)	99.8%	-0.00	0.00
Pool 06: 0.11 - 0.13	10.1%	10.0%	0.07%	101%	0.01	0.00
Pool 07: 0.09 - 0.11	9.96%	10.0%	(0.04%)	99.6%	-0.00	0.00
Pool 08: 0.06 - 0.09	10.0%	10.0%	0.01%	100%	0.00	0.00
Pool 09: 0.04 - 0.06	9.96%	10.0%	(0.04%)	99.6%	-0.00	0.00
Pool 10: 0.00 - 0.04	10.0%	10.0%	(0.00%)	100%	-0.00	0.00

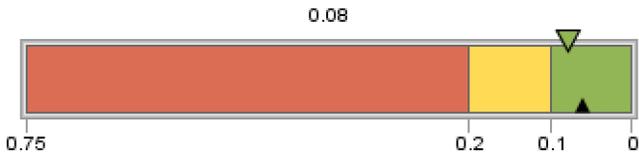
**Notes:**

Economic climates and business environments change. These changes can influence population distributions and affect the performance of a scoring system. To determine if a scoring system can continue to be used effectively, the stability of the target population needs to be measured. SSI measures the degree of change within the target populations, by calculating an Index. High index values indicate that the population has changed and may suggest that the scoring system needs re-calibration or re-development. Shifts in scorecard performance by score range may demand adjustments to score cut-offs.

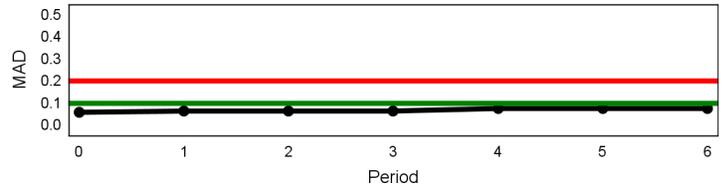
**Model validation report: Application-level loss given default  
LGD model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Error measures analysis: MAD, MSE and MAPE**

**Mean absolute deviation (MAD) benchmarks**



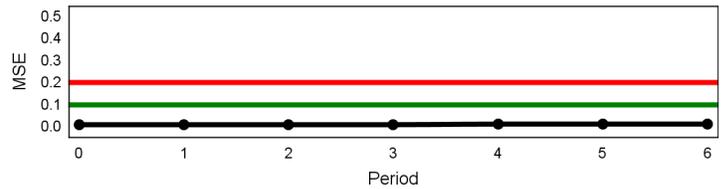
**Mean absolute deviation (MAD) trend**



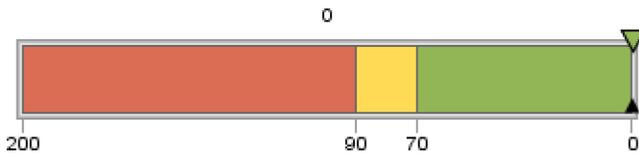
**Mean squared error (MSE) benchmarks**



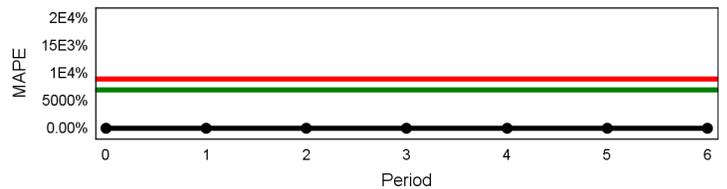
**Mean squared error (MSE) trend**



**Mean absolute percent error (MAPE) benchmarks**

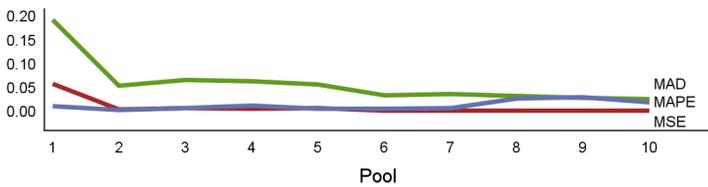


**Mean absolute percent error (MAPE) trend**

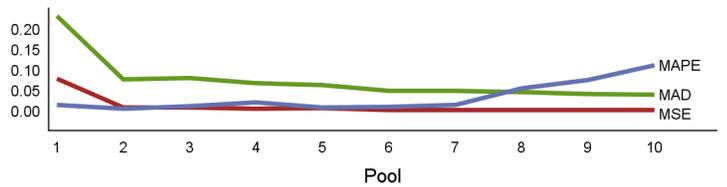


**Comparison of error measures for development and current periods**

**Error measures by pool, Period 0**



**Error measures by pool, Period 6**



**Notes:**

- Mean absolute deviation is the distance between account-level estimated and actual LGD, averaged at the pool level.
- Mean squared error is the squared distance between account-level estimated and actual LGD, averaged at the pool level.
- Mean absolute percent error is the absolute value of the account-level difference between estimated and actual LGD divided by the estimated LGD, averaged at the pool level.

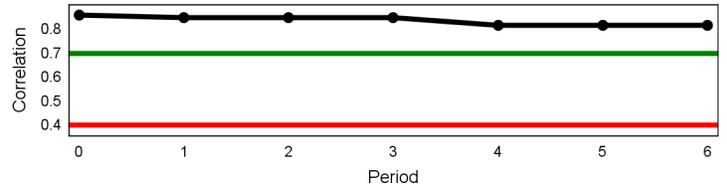
Model validation report: Application-level loss given default  
 LGD model developed on 01JAN09  
 Mortgage product scoring performance as of 01SEP10

Correlation and confidence interval analysis

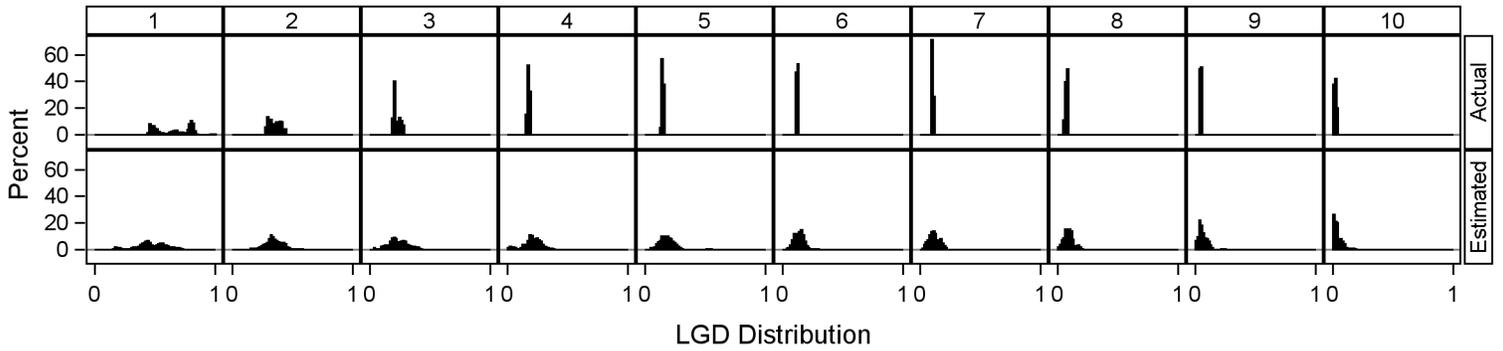
Correlation benchmarks



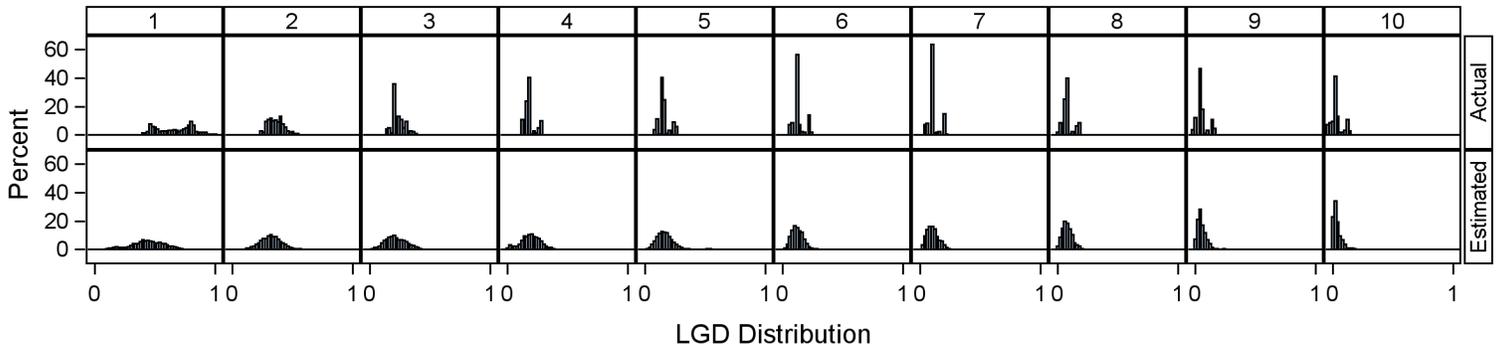
Correlation trend



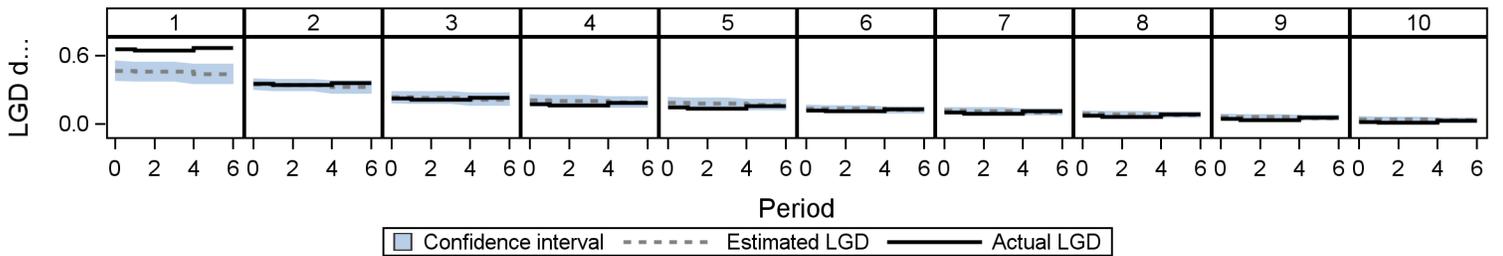
Actual vs Estimated LGD, Period 0



Actual vs Estimated LGD, Period 6



Trend in actual LGD compared with estimated confidence intervals



Notes:

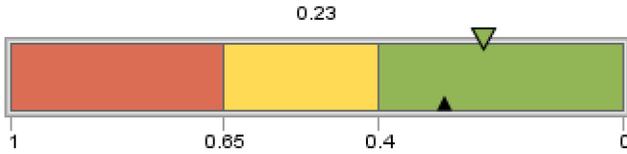
Correlation is the Pearson correlation coefficient between average and estimated LGD calculated at the pooled level.

Confidence intervals are based on pool-level average estimated LGD plus/minus the pool-level standard deviation multiplied by the 1-(alpha/2) quantile of the standard normal distribution.

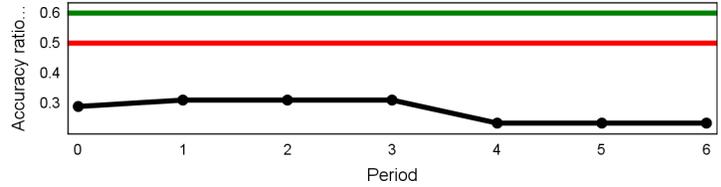
**Model validation report: Application-level loss given default  
LGD model developed on 01JAN09  
Mortgage product scoring performance as of 01SEP10**

**Accuracy ratio (Gini) analysis**

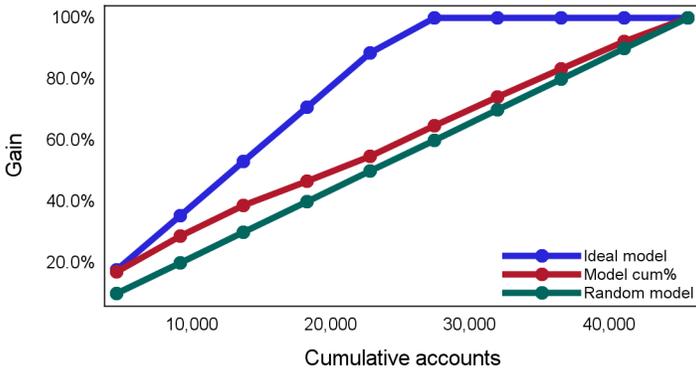
**Accuracy ratio (Gini) benchmarks**



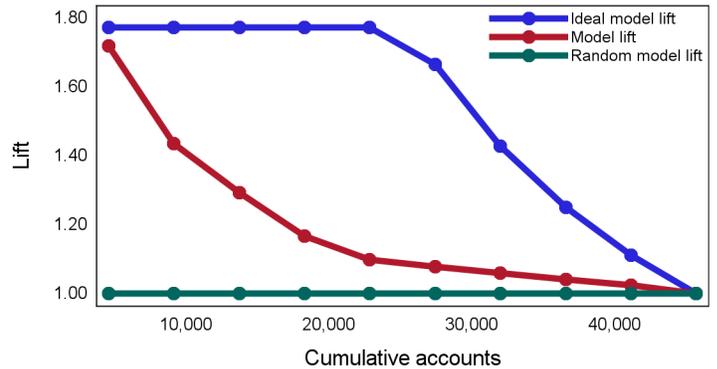
**AR (Gini) trend**



**Model gain**



**Model lift**



**Accuracy ratio (Gini) metrics by pool**

pool_id	Random model	Model cum%	Ideal model	Cumulative accounts across pools
Pool 01: 0.45 - 1.00	10.0%	17.2%	17.8%	4,582
Pool 02: 0.28 - 0.45	20.0%	28.8%	35.5%	9,155
Pool 03: 0.19 - 0.28	30.0%	38.8%	53.2%	13,717
Pool 04: 0.16 - 0.19	40.0%	46.7%	70.9%	18,281
Pool 05: 0.13 - 0.16	50.0%	54.9%	88.6%	22,839
Pool 06: 0.11 - 0.13	60.1%	64.7%	100%	27,438
Pool 07: 0.09 - 0.11	70.0%	74.2%	100%	31,986
Pool 08: 0.06 - 0.09	80.0%	83.3%	100%	36,556
Pool 09: 0.04 - 0.06	90.0%	92.2%	100%	41,107
Pool 10: 0.00 - 0.04	100%	100%	100%	45,673

**Notes:**

This analysis examines LGD events, defined as the proportion of cases per pool where actual LGD exceeds estimated LGD.

The gain curve is also known as the Gini curve, Power curve, or Lorenz curve. The gain curve is drawn by taking the cumulative percentage of all accounts (in descending order of LGD volume) on the horizontal axis and the cumulative percentage of all LGD events on the vertical axis. The higher the gain curve, the higher the relative proportion of LGD events per pool.

The Accuracy Ratio (AR) is the summary index of CAP and is also known as Gini coefficient. It shows the performance of the model being evaluated by depicting the percentage of LGD events captured by the model across different scores. The higher the AR curve, the higher the relative proportion of LGD events per pool.

Lift charts consist of a lift curve and a baseline. The baseline reflects the effectiveness when no model is used and the lift curve reflects the effectiveness when the predictive model is used. Lift is a measure of the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model. Typically, greater area between the lift curve and the baseline indicates a better model. In this case, because this analysis examines LGD events, the greater the area between the lift curve and the baseline indicates a higher incidence of LGD events, and hence a model that is more likely to under-estimate the actual number of such events.