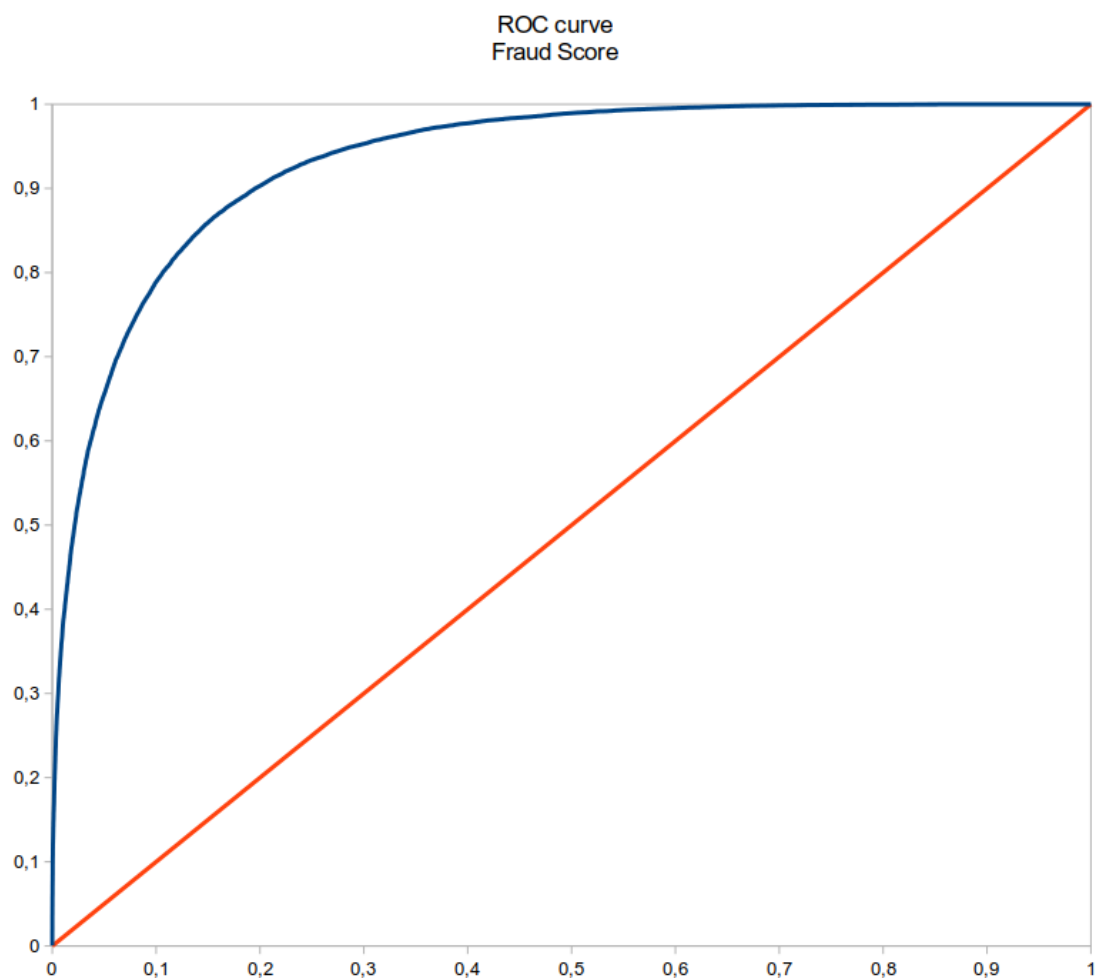


Отчет по тестированию модели ИДО (Индекс должной осмотрительности)

Набор данных для разработки был разделен на обучающую (70%) и тестовую (30%) выборки.

Для тестовой выборки:

- AUC = 0,9336.
- Коэффициент Джини = 0,8673.
- Положительная прогностическая ценность (PPV) = 74,11%.
- Полнота (доля истинно-положительных результатов) = 86,73%.
- Специфичность (доля истинно-отрицательных результатов) = 84,26%.
- Отрицательная прогностическая ценность (NPV) = 92,44%.
- Точность модели = 85,11%.
- F1-score = 0,7993.



Значение KS (тест Колмогорова-Смирнова) для обучающей выборки = 0,7067:

Bucket	Доля bad (%)	Доля good (%)	Кумулятивная доля bad (%)	Кумулятивная доля good (%)	KS
1	28,46%	0,52%	28,46%	0,52%	27,94%
2	25,26%	2,16%	53,72%	2,67%	51,05%
3	19,86%	4,93%	73,59%	7,61%	65,98%
4	13,10%	8,41%	86,68%	16,01%	70,67%
5	7,52%	11,27%	94,20%	27,29%	66,92%
6	3,64%	13,27%	97,84%	40,56%	57,29%
7	1,51%	14,36%	99,35%	54,92%	44,43%

Bucket	Доля bad (%)	Доля good (%)	Кумулятивная доля bad (%)	Кумулятивная доля good (%)	KS
8	0,51%	14,87%	99,86%	69,79%	30,07%
9	0,12%	15,08%	99,98%	84,87%	15,11%
10	0,02%	15,13%	100,00%	100,00%	0,00%

Значение KS (тест Колмогорова-Смирнова) для тестовой выборки = 0,7100:

Bucket	Доля bad (%)	Доля good (%)	Кумулятивная доля bad (%)	Кумулятивная доля good (%)	KS
1	28,30%	0,50%	28,30%	0,50%	27,80%
2	25,17%	2,12%	53,47%	2,61%	50,86%
3	19,89%	4,86%	73,36%	7,48%	65,88%
4	13,37%	8,25%	86,73%	15,73%	71,00%
5	7,52%	11,29%	94,24%	27,01%	67,23%
6	3,55%	13,35%	97,80%	40,37%	57,43%
7	1,49%	14,42%	99,29%	54,78%	44,50%
8	0,56%	14,90%	99,85%	69,69%	30,16%
9	0,12%	15,13%	99,97%	84,82%	15,16%
10	0,03%	15,18%	100,00%	100,00%	0,00%

Рассмотрим распределение компаний в каждом интервале переменных (обучающая выборка = train, тестовая выборка = test):

№ переменной	Имя переменной	Интервалы	Доля в % (test)	Доля в % (train)	PSI
1	ФАКТ_4_woe	1	90,10%	89,93%	0,0003%
		2	2,43%	2,43%	0,0000%
		3	7,47%	7,63%	0,0037%
2	ФАКТ_6_woe	1	55,63%	56,11%	0,0042%
		2	25,94%	25,51%	0,0074%
		3	13,58%	13,63%	0,0002%
		4	4,84%	4,74%	0,0020%
3	ФАКТ_8_woe	1	55,64%	55,82%	0,0006%
		2	26,56%	26,50%	0,0001%
		3	13,43%	13,40%	0,0001%
		4	4,37%	4,28%	0,0016%
4	ФАКТ_9_12_woe	1	11,90%	11,88%	0,0001%
		2	88,10%	88,12%	0,0000%
5	ФАКТ_15_woe	1	9,66%	9,68%	0,0001%
		2	90,34%	90,32%	0,0000%
6	ФАКТ_17_woe	1	74,17%	74,20%	0,0000%
		2	18,13%	18,27%	0,0010%
		3	7,70%	7,54%	0,0033%
7	ФАКТ_19_woe	1	12,68%	12,67%	0,0000%
		2	7,67%	7,71%	0,0002%
		3	11,81%	11,78%	0,0001%
		4	67,85%	67,84%	0,0000%
8	ФАКТ_21_1_woe	1	21,58%	21,54%	0,0001%
		2	78,42%	78,46%	0,0000%
9	ФАКТ_22_woe	1	3,44%	3,45%	0,0000%
		2	0,87%	0,86%	0,0000%
		3	9,19%	9,26%	0,0005%
		4	78,20%	78,41%	0,0005%
		5	8,29%	8,02%	0,0092%
10	ФАКТ_24_woe	1	11,78%	11,74%	0,0001%
		2	88,22%	88,26%	0,0000%
11	ФАКТ_47_woe	1	56,46%	56,48%	0,0000%
		2	43,54%	43,52%	0,0000%
12	ФАКТ_48_woe	1	54,90%	54,70%	0,0008%

№ переменной	Имя переменной	Интервалы	Доля в % (test)	Доля в % (train)	PSI
		2	45,10%	45,30%	0,0009%
13	FACT_49_woe	1	6,35%	6,33%	0,0001%
		2	93,65%	93,67%	0,0000%
14	FACT_62_woe	1	41,31%	41,19%	0,0003%
		2	58,69%	58,81%	0,0002%
15	FACT_66_woe	1	83,98%	84,08%	0,0001%
		2	16,02%	15,92%	0,0006%
16	FACT_67_woe	1	18,05%	18,22%	0,0016%
		2	81,95%	81,78%	0,0004%
17	FACT_68_woe	1	15,31%	15,40%	0,0005%
		2	84,69%	84,60%	0,0001%
18	FACT_69_woe	1	48,39%	48,53%	0,0004%
		2	29,80%	29,80%	0,0000%
		3	15,54%	15,44%	0,0006%
		4	6,27%	6,23%	0,0002%
19	STAFF_woe	1	34,80%	35,01%	0,0012%
		2	43,70%	43,67%	0,0000%
		3	21,50%	21,32%	0,0014%
20	FACT_378_woe	1	72,68%	72,93%	0,0009%
		2	14,21%	14,18%	0,0001%
		3	13,11%	12,90%	0,0036%
21	FACT_379_woe	1	74,48%	74,50%	0,0000%
		2	12,62%	12,65%	0,0000%
		3	12,89%	12,85%	0,0001%
22	FACT_380_woe	1	20,54%	20,46%	0,0004%
		2	79,46%	79,54%	0,0001%
23	FACT_381_9_woe	1	6,86%	6,82%	0,0002%
		2	4,54%	4,59%	0,0005%
		3	16,38%	16,16%	0,0028%
		4	72,22%	72,43%	0,0006%
24	FACT_381_13_woe	1	46,46%	46,78%	0,0021%
		2	36,47%	36,20%	0,0021%
		3	10,98%	11,01%	0,0001%
		4	6,08%	6,02%	0,0007%
25	FACT_82_1_woe	1	8,81%	8,82%	0,0000%
		2	91,19%	91,18%	0,0000%
26	FACT_82_2_woe	1	22,66%	23,00%	0,0048%
		2	77,34%	77,00%	0,0014%
27	FACT_83_woe	1	15,25%	15,44%	0,0024%
		2	84,75%	84,56%	0,0004%
28	FACT_85_woe	1	16,15%	16,00%	0,0014%
		2	83,85%	84,00%	0,0003%
29	FACT_87_woe	1	16,78%	16,73%	0,0002%
		2	83,22%	83,27%	0,0000%
30	FACT_88_woe	1	22,05%	21,96%	0,0004%
		2	77,95%	78,04%	0,0001%
31	FACT_89_woe	1	20,06%	20,28%	0,0023%
		2	79,94%	79,72%	0,0006%
32	FACT_90_91_1_woe	1	77,96%	77,94%	0,0000%
		2	22,04%	22,06%	0,0000%
33	FACT_90_91_3_woe	1	37,03%	37,33%	0,0024%
		2	62,97%	62,67%	0,0014%
34	FACT_90_91_4_woe	1	77,02%	77,18%	0,0003%

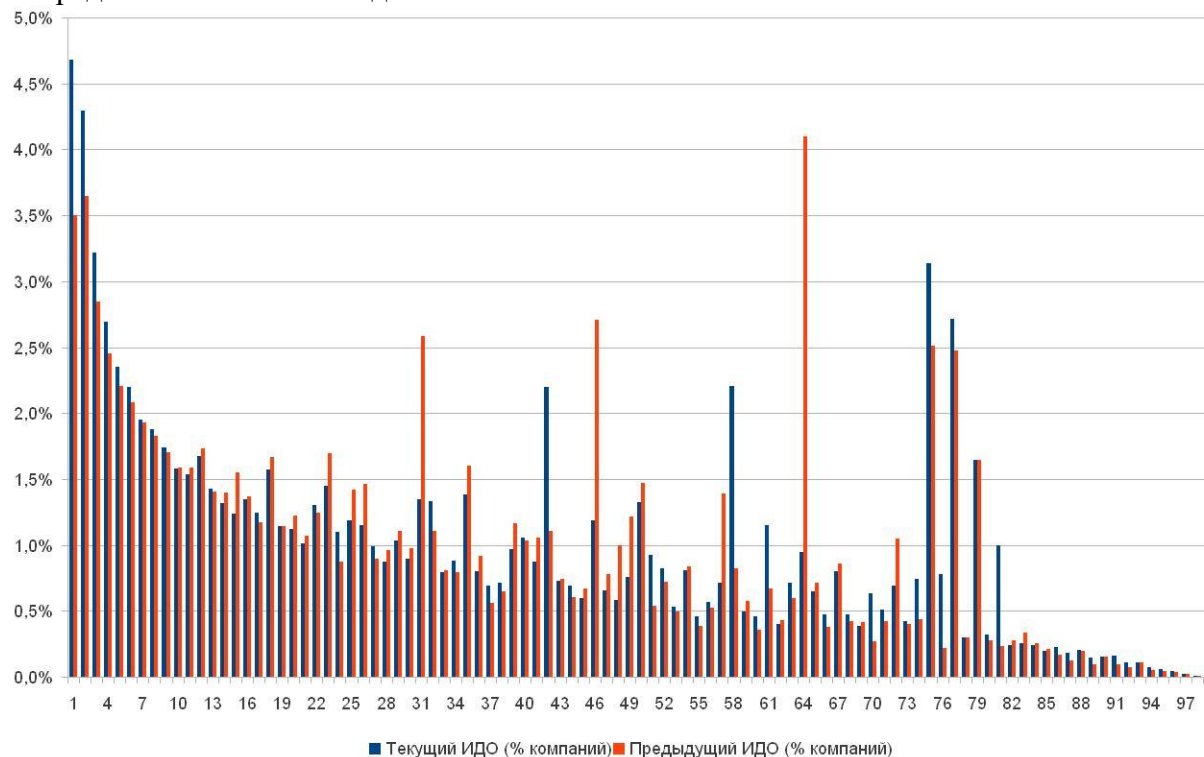
№ переменной	Имя переменной	Интервалы	Доля в % (test)	Доля в % (train)	PSI
		2	22,98%	22,82%	0,0011%
35	FACT_90_91_6_woe	1	84,70%	84,72%	0,0000%
		2	15,30%	15,28%	0,0000%
36	CRITICAL_SUM_woe	1	67,51%	67,38%	0,0003%
		2	17,61%	17,66%	0,0002%
		3	11,10%	11,08%	0,0000%
		4	2,76%	2,82%	0,0009%
		5	1,02%	1,07%	0,0021%
37	FACT_97_1_woe	1	8,05%	8,12%	0,0007%
		2	91,95%	91,88%	0,0001%
38	FACT_97_2_woe	1	9,67%	9,71%	0,0001%
		2	90,33%	90,29%	0,0000%
39	FACT_98_woe	1	2,39%	2,39%	0,0000%
		2	4,70%	4,66%	0,0004%
		3	6,02%	5,98%	0,0002%
		4	86,90%	86,97%	0,0001%
40	FACT_99_1_woe	1	8,04%	8,05%	0,0000%
		2	11,06%	11,08%	0,0000%
		3	13,11%	13,29%	0,0024%
		4	16,47%	16,40%	0,0003%
		5	51,32%	51,18%	0,0004%
41	FACT_100_woe	1	31,14%	31,28%	0,0006%
		2	68,86%	68,72%	0,0003%
42	FACT_107_woe	1	2,40%	2,48%	0,0026%
		2	4,38%	4,45%	0,0011%
		3	4,03%	4,15%	0,0034%
		4	6,83%	6,82%	0,0000%
		5	82,36%	82,11%	0,0008%
43	FACT_108_woe	1	6,15%	6,24%	0,0013%
		2	93,85%	93,76%	0,0001%
44	IS_ACTIVE_woe	1	18,57%	18,48%	0,0004%
		2	81,43%	81,52%	0,0001%
45	NO_CHECKS_woe	1	76,60%	76,84%	0,0008%
		2	6,14%	6,10%	0,0003%
		3	5,88%	5,85%	0,0001%
		4	5,75%	5,64%	0,0022%
		5	5,63%	5,57%	0,0007%
46	REVENUE_DECREASE_50PC_woe	1	59,93%	59,96%	0,0000%
		2	40,07%	40,04%	0,0000%

Устойчивость модели относительно данных

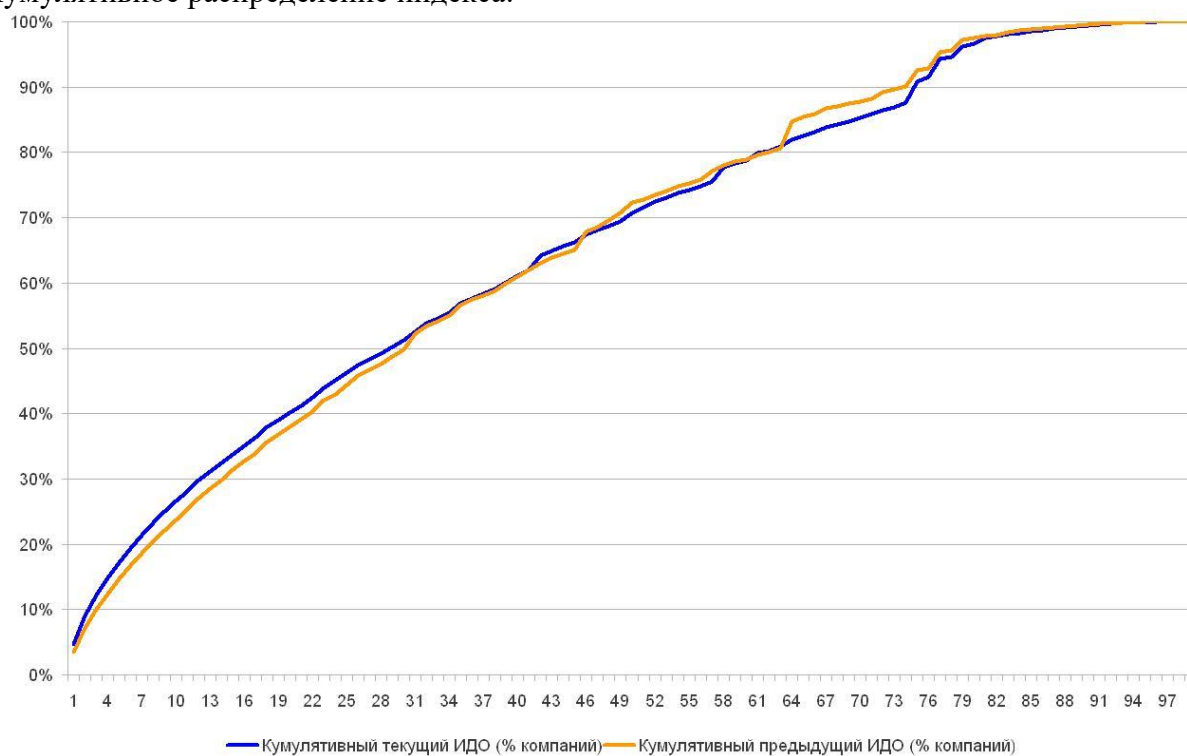
Рассмотрим, как ведет себя текущая модель (до обновления) на двух генеральных совокупностях:

- 1) в момент построения;
- 2) в текущий момент.

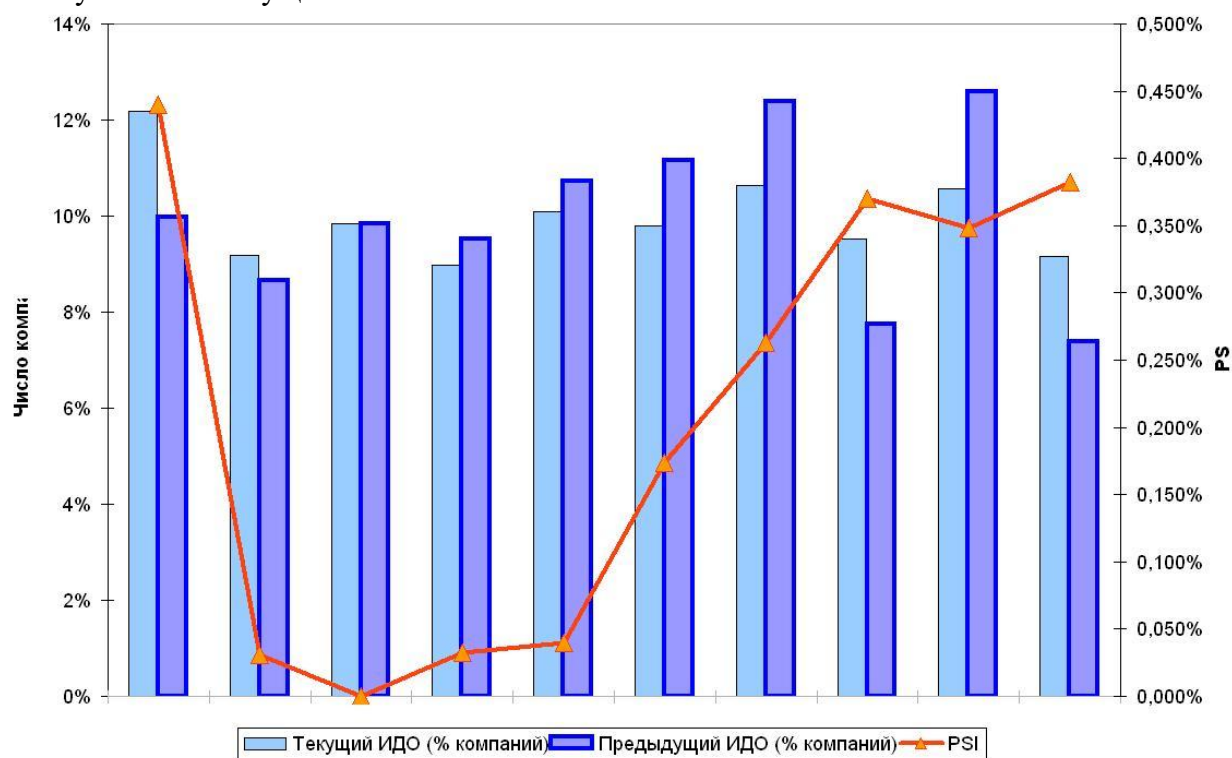
Распределение значений индекса:



Кумулятивное распределение индекса:

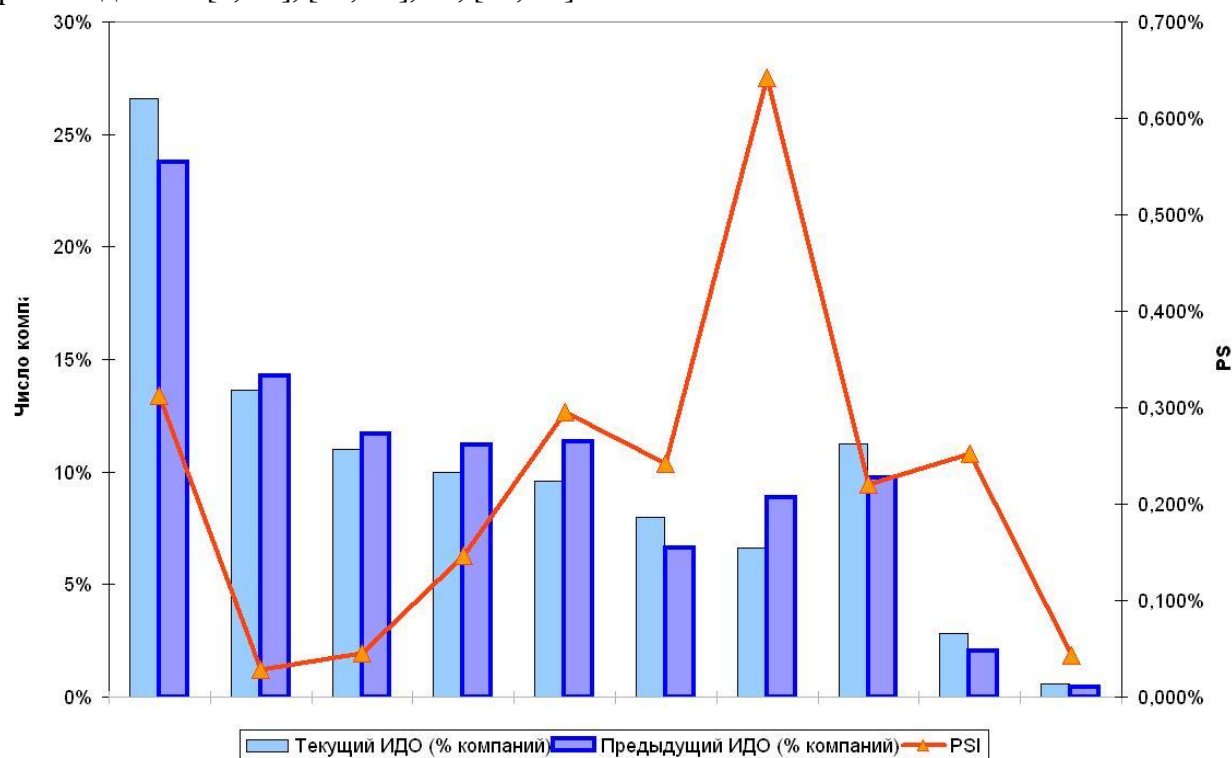


Распределение квантилей – значения индекса разбиты на 10 групп таким образом, чтобы в каждой из них находилось примерно равно число наблюдений для генеральной совокупности в текущий момент:



Итоговый PSI = 2,079%.

Аналогично предыдущему распределению, но значения индекса разбиты на 10 групп равной длины: [1, 10], [11, 20], ..., [91, 99]:



Итоговый PSI = 2,231%.

Рассмотрим устойчивость отдельных переменных:

№ переменной	Имя переменной	PSI
1	CRITICAL_SUM	0,535%
2	FACT_100	0,025%
3	FACT_107	0,017%
4	FACT_108	0,001%
5	FACT_15	0,000%
6	FACT_17	0,157%
7	FACT_19	0,004%
8	FACT_21_1	0,003%
9	FACT_22	0,032%
10	FACT_24	0,000%
11	FACT_378	0,000%
12	FACT_379	0,000%
13	FACT_380	0,000%
14	FACT_381_13	0,036%
15	FACT_381_9	0,029%
16	FACT_4	0,299%
17	FACT_47	0,002%
18	FACT_48	0,001%
19	FACT_49	0,361%
20	FACT_6	0,136%
21	FACT_62	0,003%
22	FACT_66	0,004%
23	FACT_67	0,012%
24	FACT_68	0,020%
25	FACT_69	0,001%
26	FACT_8	0,005%
27	FACT_82_1	0,013%
28	FACT_82_2	0,001%
29	FACT_83	0,027%
30	FACT_85	0,001%
31	FACT_87	0,000%
32	FACT_88	0,000%
33	FACT_89	0,001%
34	FACT_9_12	0,056%
35	FACT_90_91_1	0,000%
36	FACT_90_91_3	0,003%
37	FACT_90_91_4	0,000%
38	FACT_90_91_6	0,000%
39	FACT_97_1	0,005%
40	FACT_97_2	0,001%
41	FACT_98	0,006%
42	FACT_99_1	0,006%
43	IS_ACTIVE	0,036%
44	NO_CHECKS	0,318%
45	REVENUE_DECREASE_50PC	0,054%
46	STAFF	0,020%

Рассмотрим число пропусков в обеих генеральных совокупностях:

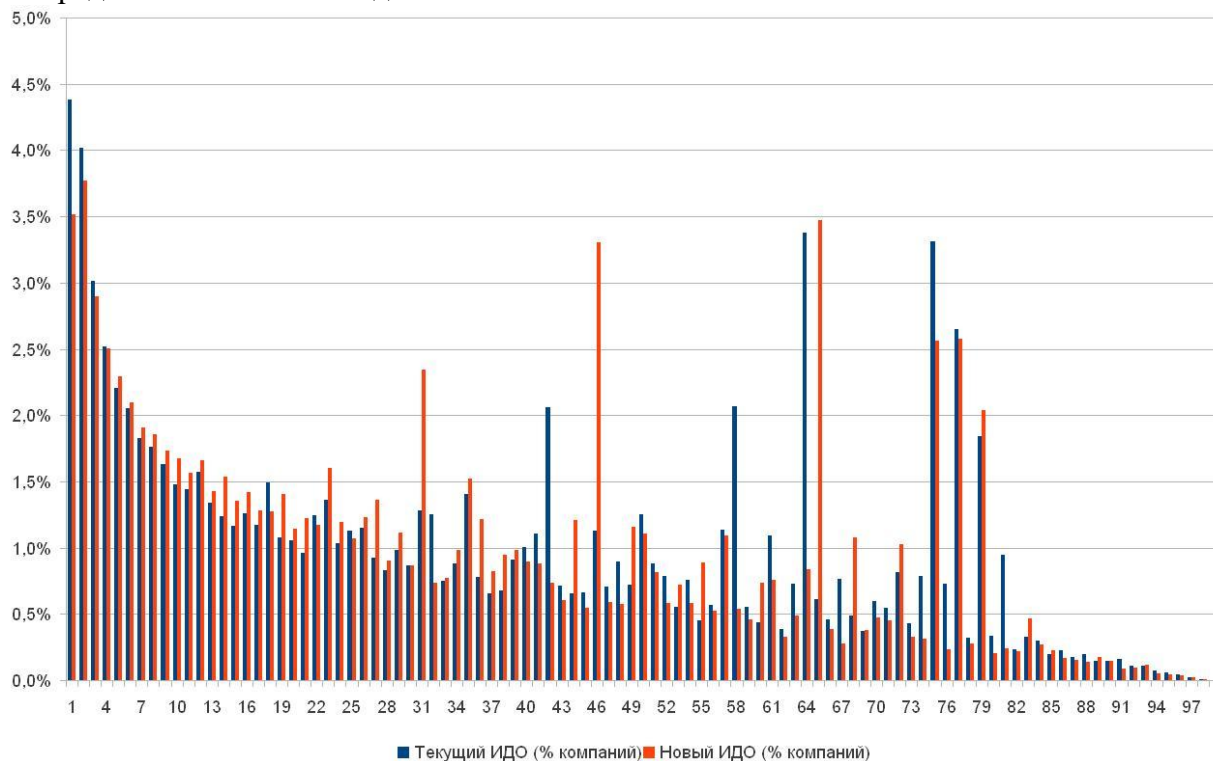
№ переменной	Имя переменной	Новая ген совокупность		Предыдущая ген совокупность	
		Пропуски	% пропусков	Пропуски	% пропусков
1	CRITICAL_SUM	425 732	16,3%	442 951	17,1%
2	FACT_100	0	0,0%	0	0,0%
3	FACT_107	2 358 429	90,6%	2 346 707	90,5%

№ переменной	Имя переменной	Новая ген совокупность		Предыдущая ген совокупность	
		Пропуски	% пропусков	Пропуски	% пропусков
4	FACT_108	0	0,0%	0	0,0%
5	FACT_15	0	0,0%	0	0,0%
6	FACT_17	0	0,0%	0	0,0%
7	FACT_19	9 971	0,4%	10 983	0,4%
8	FACT_21_1	0	0,0%	0	0,0%
9	FACT_22	0	0,0%	0	0,0%
10	FACT_24	0	0,0%	0	0,0%
11	FACT_378	0	0,0%	0	0,0%
12	FACT_379	0	0,0%	0	0,0%
13	FACT_380	0	0,0%	0	0,0%
14	FACT_381_13	0	0,0%	0	0,0%
15	FACT_381_9	0	0,0%	0	0,0%
16	FACT_4	0	0,0%	0	0,0%
17	FACT_47	399 260	15,3%	443 331	17,1%
18	FACT_48	61	0,0%	143	0,0%
19	FACT_49	0	0,0%	0	0,0%
20	FACT_6	0	0,0%	0	0,0%
21	FACT_62	0	0,0%	0	0,0%
22	FACT_66	0	0,0%	0	0,0%
23	FACT_67	0	0,0%	0	0,0%
24	FACT_68	0	0,0%	0	0,0%
25	FACT_69	0	0,0%	0	0,0%
26	FACT_8	0	0,0%	0	0,0%
27	FACT_82_1	0	0,0%	0	0,0%
28	FACT_82_2	0	0,0%	0	0,0%
29	FACT_83	0	0,0%	0	0,0%
30	FACT_85	0	0,0%	0	0,0%
31	FACT_87	0	0,0%	0	0,0%
32	FACT_88	0	0,0%	0	0,0%
33	FACT_89	0	0,0%	0	0,0%
34	FACT_9_12	0	0,0%	0	0,0%
35	FACT_90_91_1	0	0,0%	0	0,0%
36	FACT_90_91_3	0	0,0%	0	0,0%
37	FACT_90_91_4	0	0,0%	0	0,0%
38	FACT_90_91_6	0	0,0%	0	0,0%
39	FACT_97_1	669 101	25,7%	670 237	25,9%
40	FACT_97_2	669 101	25,7%	670 237	25,9%
41	FACT_98	2 434 501	93,5%	2 418 089	93,3%
42	FACT_99_1	2 434 189	93,5%	2 421 341	93,4%
43	IS_ACTIVE	0	0,0%	0	0,0%
44	NO_CHECKS	2 491 386	95,7%	2 482 537	95,8%
45	REVENUE_DECREASE_50PC	2 491 386	95,7%	2 482 537	95,8%
46	STAFF	550 314	21,1%	543 177	21,0%

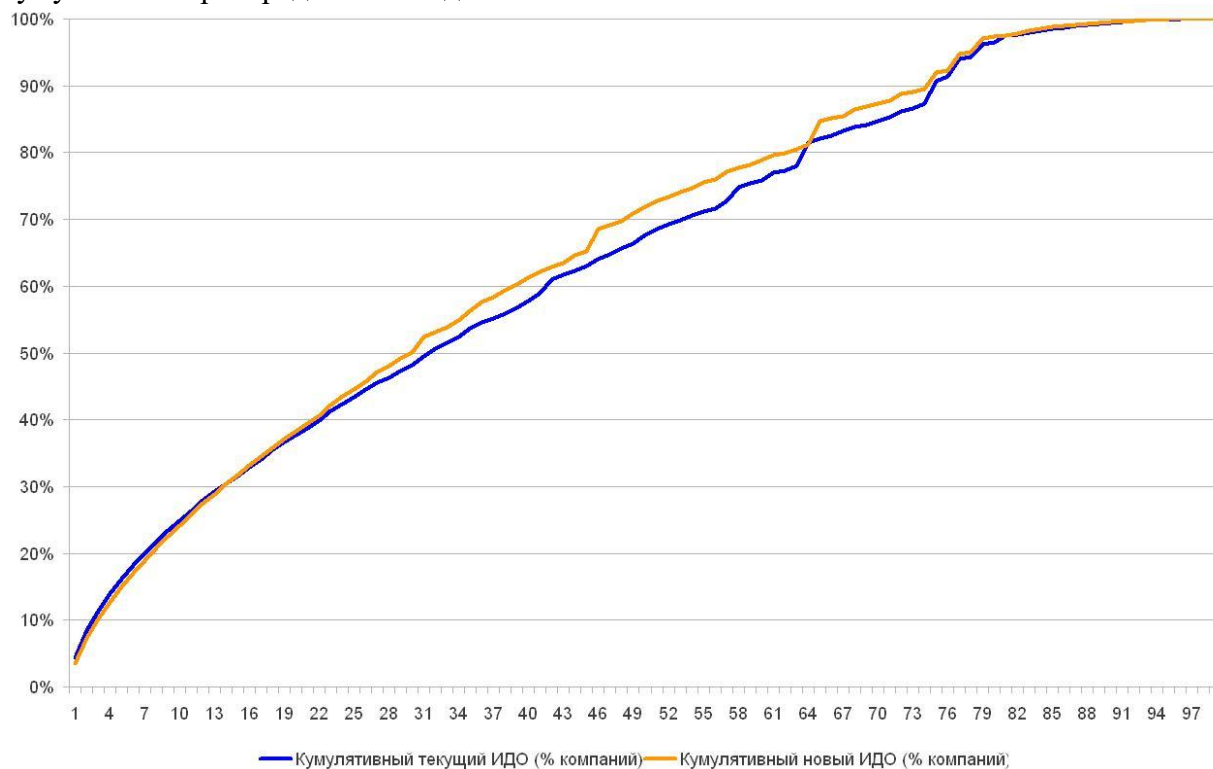
Устойчивость модели относительно версий

Рассмотрим устойчивость модели относительно процедуры обновления на генеральной совокупности в текущий момент.

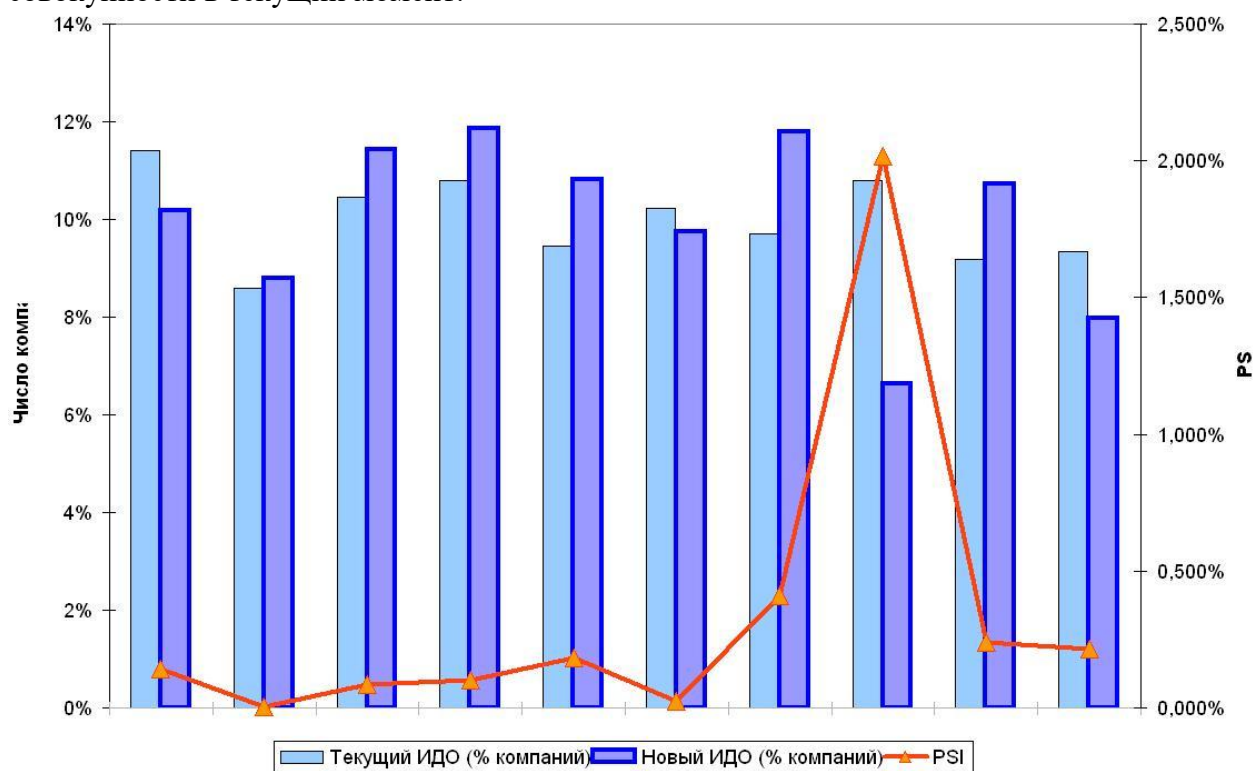
Распределение значений индекса:



Кумулятивное распределение индекса:

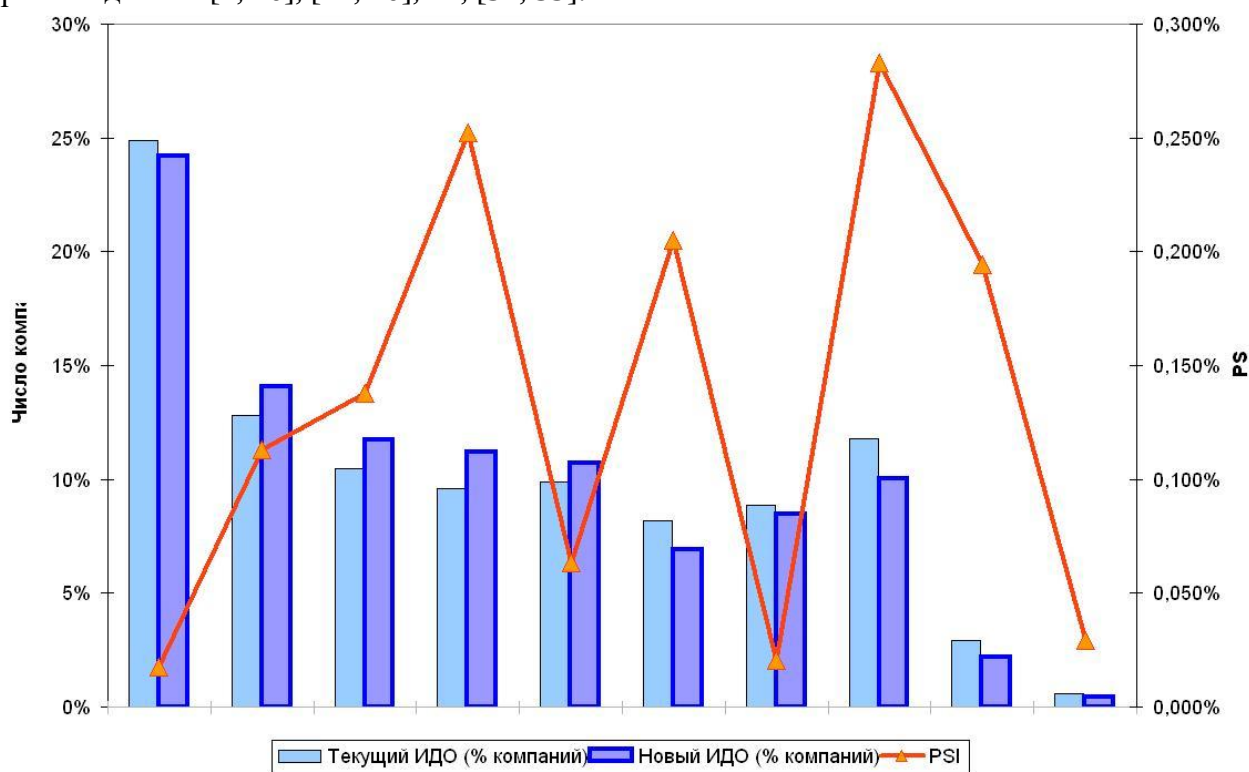


Распределение квантилей – значения индекса разбиты на 10 групп таким образом, чтобы в каждой из них находилось примерно равно число наблюдений для генеральной совокупности в текущий момент:



Итоговый PSI = 3,422%.

Аналогично предыдущему распределению, но значения индекса разбиты на 10 групп равной длины: [1, 10], [11, 20], ..., [91, 99]:



Итоговый PSI = 1,316%.