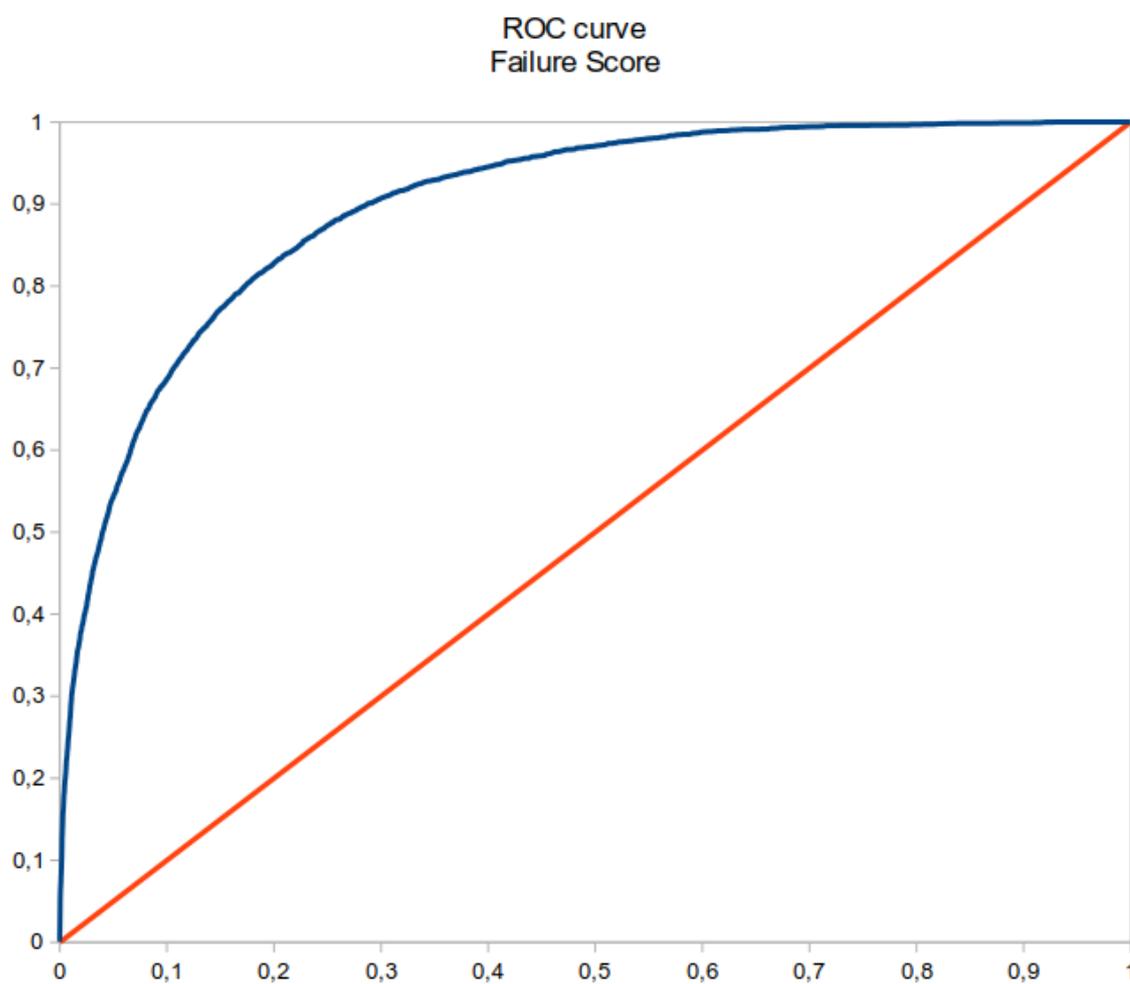


Отчет по тестированию модели ИФР (Индекс финансового риска)

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

Для тестовой выборки:

- AUC = 0,8977.
- Коэффициент Джини = 0,7953.
- Положительная прогностическая ценность (PPV) = 74,63%.
- Полнота (доля истинно-положительных результатов) = 81,64%.
- Специфичность (доля истинно-отрицательных результатов) = 80,53%.
- Отрицательная прогностическая ценность (NPV) = 86,22%.
- Точность модели = 80,99%.
- F1-score = 0,7798.



Значение KS (тест Колмогорова-Смирнова) для обучающей выборки = 0,6284:

Bucket	Доля bad (%)	Доля good (%)	Кумулятивная доля bad (%)	Кумулятивная доля good (%)	KS
1	61,75%	7,36%	61,75%	7,36%	54,40%
2	18,03%	9,59%	79,79%	16,95%	62,84%
3	8,69%	10,07%	88,48%	27,01%	61,46%
4	4,98%	10,26%	93,46%	37,27%	56,19%
5	3,02%	10,36%	96,48%	47,63%	48,85%
6	1,60%	10,43%	98,08%	58,06%	40,02%
7	1,01%	10,46%	99,08%	68,51%	30,57%
8	0,56%	10,48%	99,64%	79,00%	20,65%

Bucket	Доля bad (%)	Доля good (%)	Кумулятивная доля bad (%)	Кумулятивная доля good (%)	KS
9	0,25%	10,50%	99,90%	89,49%	10,40%
10	0,10%	10,51%	100,00%	100,00%	0,00%

Значение KS (тест Колмогорова-Смирнова) для тестовой выборки = 0,6253:

Bucket	Доля bad (%)	Доля good (%)	Кумулятивная доля bad (%)	Кумулятивная доля good (%)	KS
1	62,28%	7,24%	62,28%	7,24%	55,04%
2	17,11%	9,62%	79,39%	16,86%	62,53%
3	9,40%	10,03%	88,79%	26,89%	61,90%
4	4,85%	10,27%	93,64%	37,17%	56,47%
5	3,01%	10,37%	96,65%	47,53%	49,11%
6	1,81%	10,43%	98,46%	57,97%	40,49%
7	0,90%	10,48%	99,36%	68,45%	30,91%
8	0,28%	10,51%	99,63%	78,96%	20,67%
9	0,25%	10,51%	99,89%	89,48%	10,41%
10	0,11%	10,52%	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	RAZDEL_woe	1	71,75%	71,67%	0,0001%
		2	6,98%	7,01%	0,0001%
		3	12,27%	12,43%	0,0019%
		4	8,99%	8,89%	0,0011%
2	REGION_woe	1	21,26%	21,19%	0,0002%
		2	25,64%	25,69%	0,0001%
		3	47,96%	47,94%	0,0000%
		4	5,14%	5,18%	0,0003%
3	ФАКТ_9_12_woe	1	14,61%	14,54%	0,0004%
		2	6,88%	6,81%	0,0005%
		3	78,51%	78,64%	0,0002%
4	ФАКТ_375_woe	1	79,50%	79,56%	0,0000%
		2	11,08%	10,96%	0,0013%
		3	5,07%	5,20%	0,0030%
		4	4,35%	4,29%	0,0010%
5	ФАКТ_376_379_woe	1	88,85%	88,81%	0,0000%
		2	5,22%	5,31%	0,0017%
		3	3,44%	3,43%	0,0000%
		4	2,49%	2,44%	0,0009%
6	ФАКТ_378_woe	1	93,68%	93,80%	0,0002%
		2	4,35%	4,23%	0,0037%
		3	0,93%	0,91%	0,0003%
		4	1,04%	1,06%	0,0006%
7	ФАКТ_93_woe	1	24,83%	24,69%	0,0008%
		2	21,27%	21,54%	0,0033%
		3	0,33%	0,31%	0,0010%
		4	22,94%	22,99%	0,0001%
		5	30,63%	30,47%	0,0008%
8	ФАКТ_381_13_woe	1	73,55%	73,41%	0,0003%
		2	15,02%	15,17%	0,0014%
		3	11,43%	11,42%	0,0000%
9	ФАКТ_1_3_woe	1	27,08%	27,10%	0,0000%
		2	15,30%	15,35%	0,0002%
		3	3,20%	3,20%	0,0000%
		4	17,42%	17,46%	0,0001%
		5	37,01%	36,88%	0,0004%
10	ФАКТ_2_5_woe	1	40,19%	40,03%	0,0006%
		2	34,36%	34,56%	0,0011%
		3	5,30%	5,35%	0,0007%
		4	12,17%	12,12%	0,0002%
		5	7,99%	7,95%	0,0003%
11	ФАКТ_3_6_woe	1	66,44%	66,56%	0,0002%
		2	2,75%	2,79%	0,0004%

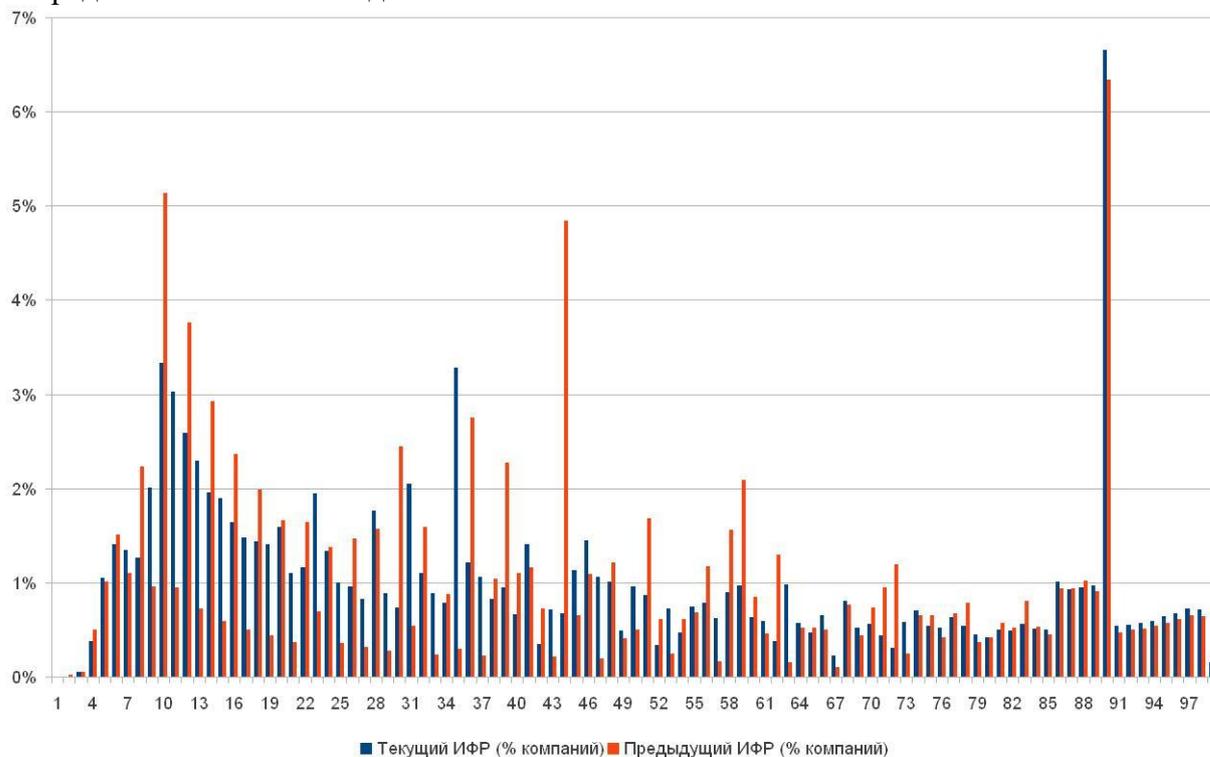
№ переменной	Имя переменной	Интервалы	Доля в % (test)	Доля в % (train)	PSI
		3	9,07%	9,39%	0,0104%
		4	7,83%	7,78%	0,0004%
		5	13,90%	13,49%	0,0122%
12	ФАКТ_3_10_woe	1	65,39%	65,06%	0,0017%
		2	0,06%	0,06%	0,0001%
		3	12,72%	13,09%	0,0107%
		4	15,00%	15,00%	0,0000%
		5	6,83%	6,79%	0,0002%
13	ФАКТ_4_4_woe	1	25,79%	26,14%	0,0048%
		2	50,16%	49,89%	0,0015%
		3	0,42%	0,45%	0,0014%
		4	13,50%	13,44%	0,0002%
		5	10,14%	10,08%	0,0003%
14	ФАКТ_4_6_woe	1	45,87%	46,09%	0,0011%
		2	0,01%	0,02%	0,0025%
		3	15,86%	16,00%	0,0011%
		4	20,30%	20,26%	0,0001%
		5	8,05%	7,76%	0,0109%
		6	9,90%	9,87%	0,0001%
15	PRBL_woe	1	36,42%	36,43%	0,0000%
		2	30,26%	30,27%	0,0000%
		3	22,63%	22,43%	0,0017%
		4	10,70%	10,86%	0,0026%
16	DENEZHSR_woe	1	54,62%	54,57%	0,0000%
		2	39,88%	39,96%	0,0002%
		3	5,50%	5,47%	0,0002%
17	SVKDOLG_woe	1	43,75%	43,55%	0,0009%
		2	56,25%	56,45%	0,0007%
18	РАВКАПИТ_woe	1	31,63%	31,52%	0,0004%
		2	33,12%	33,24%	0,0005%
		3	21,98%	21,97%	0,0000%
		4	13,26%	13,26%	0,0000%

Устойчивость модели относительно данных

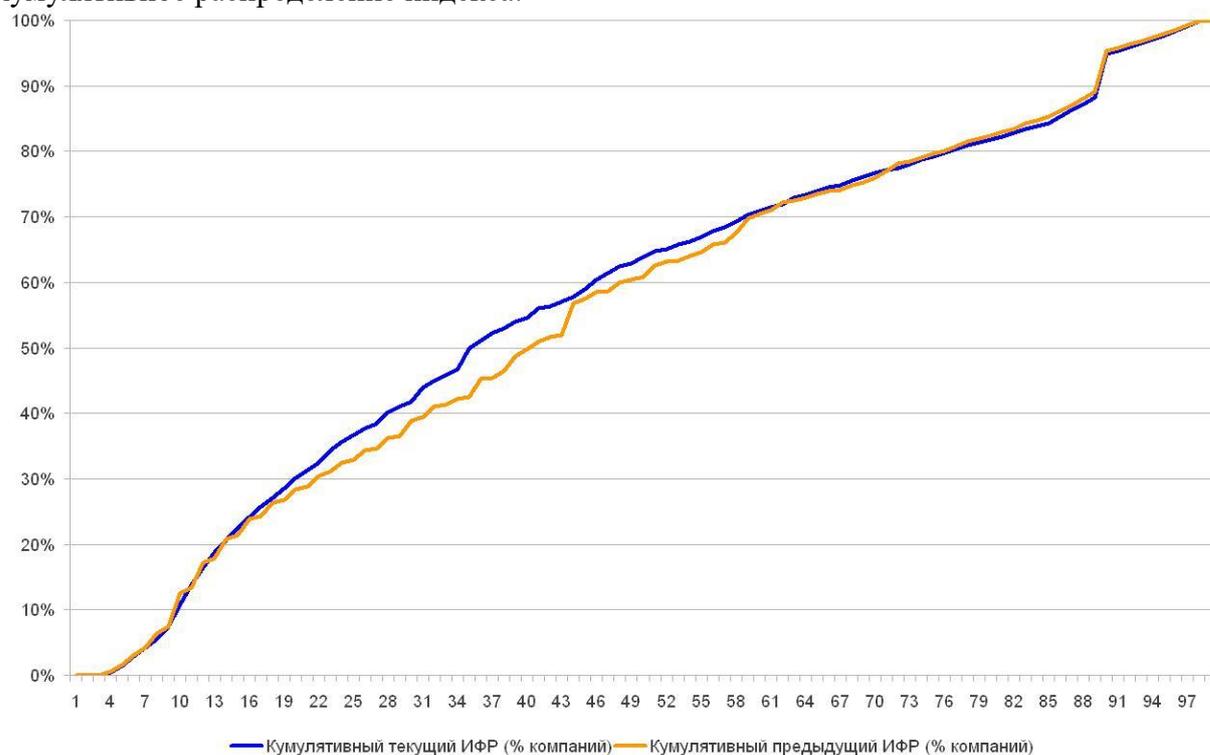
Рассмотрим, как ведет себя текущая модель (до обновления) на двух генеральных совокупностях:

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

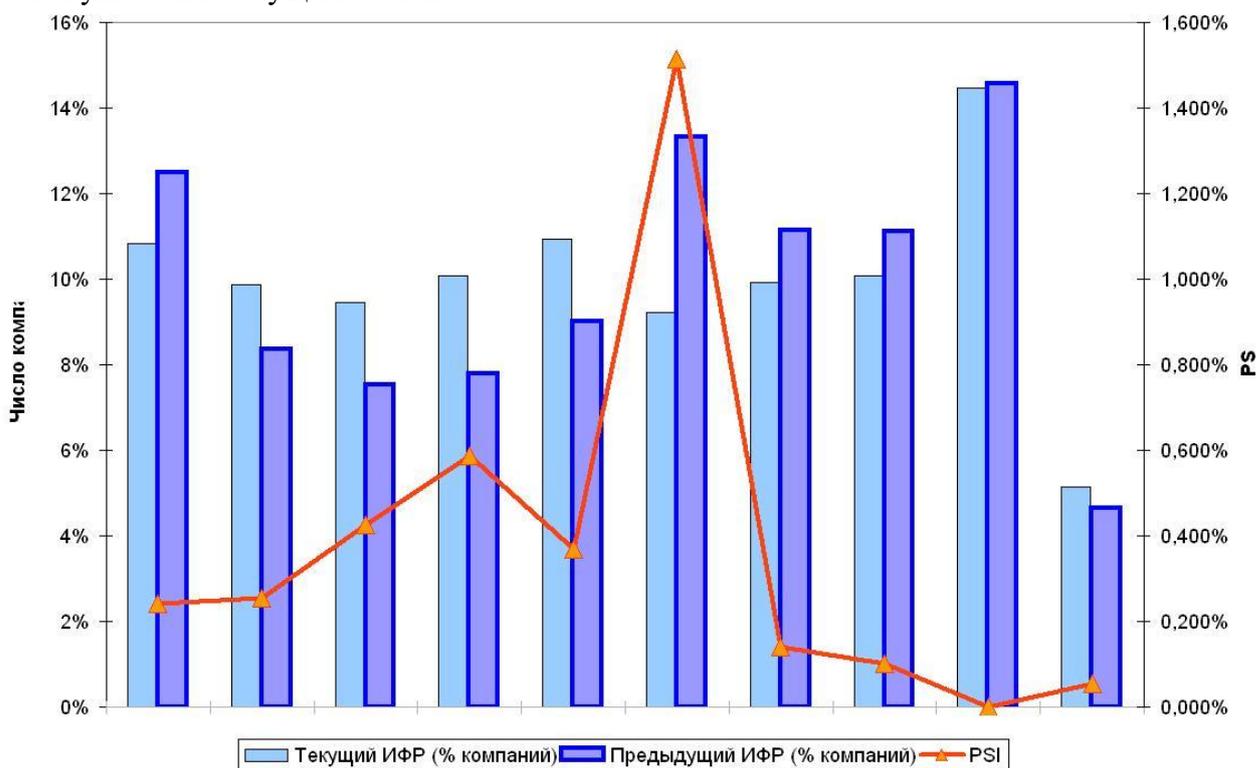
Распределение значений индекса:



Кумулятивное распределение индекса:

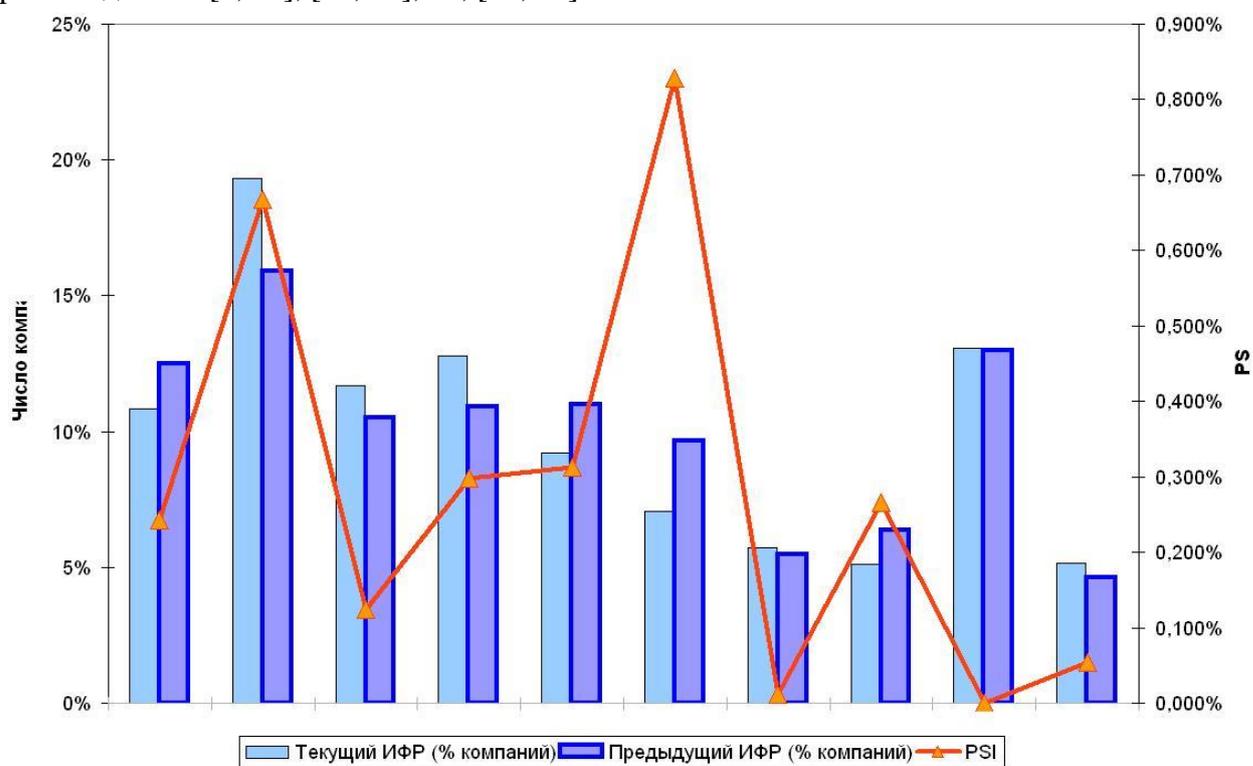


Распределение квантилей – значения индекса разбиты на 10 групп таким образом, чтобы в каждой из них находилось примерно равно число наблюдений для генеральной совокупности в текущий момент:



Итоговый PSI = 3,690%.

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



Итоговый PSI = 2,807%.

Рассмотрим устойчивость отдельных переменных:

№ переменной	Имя переменной	PSI
1	RAZDEL_woe	10,693%
2	REGION_woe	3,021%
3	FACT_9_12_woe	0,055%
4	FACT_375_woe	0,246%
5	FACT_376_379_woe	0,000%
6	FACT_378_woe	0,000%
7	FACT_93_woe	0,019%
8	FACT_381_13_woe	0,055%
9	FACT_1_3_woe	0,005%
10	FACT_2_5_woe	0,016%
11	FACT_3_6_woe	0,108%
12	FACT_3_10_woe	0,028%
13	FACT_4_4_woe	0,031%
14	FACT_4_6_woe	0,021%
15	PRBL_woe	0,059%
16	DENEZHSR_woe	0,052%
17	SVKDOLG_woe	0,418%
18	RABKAPIT_woe	0,069%

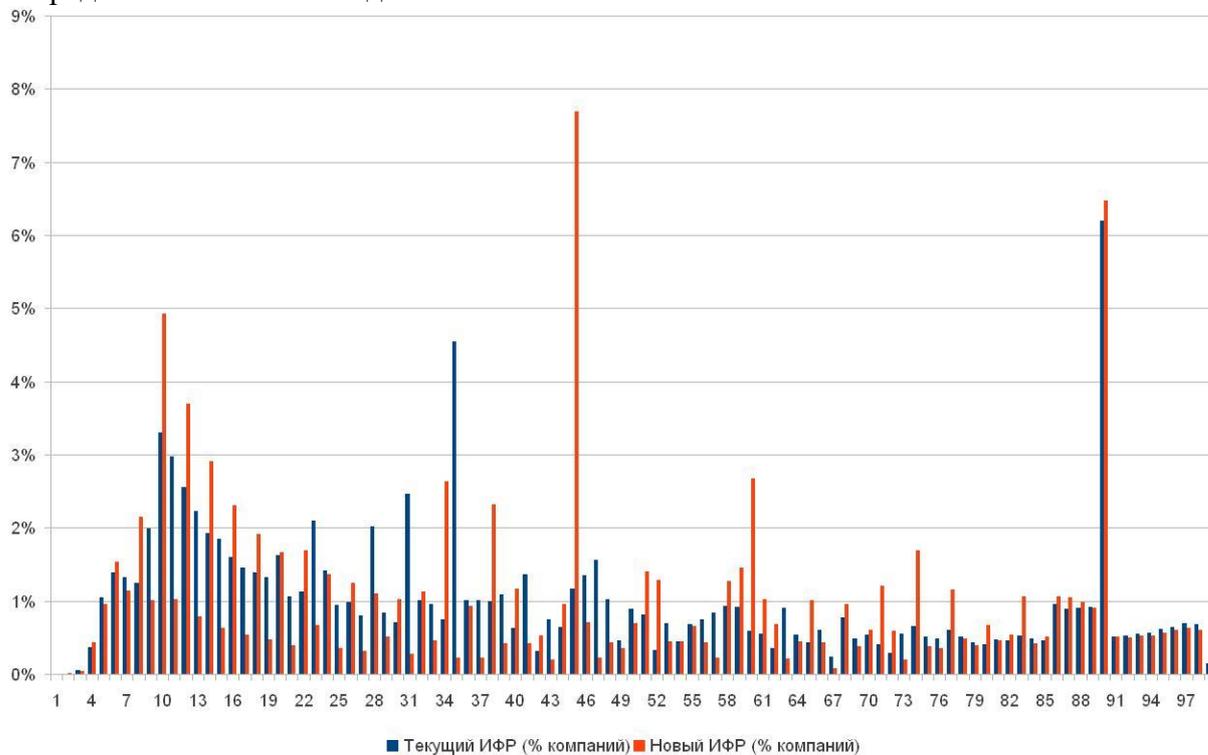
Рассмотрим число пропусков в обеих генеральных совокупностях:

№ переменной	Имя переменной	Новая ген совокупность		Предыдущая ген совокупность	
		Пропуски	% пропусков	Пропуски	% пропусков
1	RAZDEL_woe	7 793	0,30%	8 825	0,34%
2	REGION_woe	0	0,00%	10	0,00%
3	FACT_9_12_woe	0	0,00%	0	0,00%
4	FACT_375_woe	0	0,00%	0	0,00%
5	FACT_376_379_woe	0	0,00%	0	0,00%
6	FACT_378_woe	0	0,00%	0	0,00%
7	FACT_93_woe	532 589	20,72%	527 567	20,61%
8	FACT_381_13_woe	0	0,00%	0	0,00%
9	FACT_1_3_woe	971 090	37,77%	974 191	38,05%
10	FACT_2_5_woe	1 175 094	45,71%	1 183 916	46,24%
11	FACT_3_6_woe	936 422	36,43%	941 629	36,78%
12	FACT_3_10_woe	737 589	28,69%	741 163	28,95%
13	FACT_4_4_woe	1 022 735	39,78%	1 026 240	40,08%
14	FACT_4_6_woe	723 515	28,14%	726 746	28,39%
15	PRBL_woe	703 967	27,38%	706 999	27,61%
16	DENEZHSR_woe	705 067	27,43%	707 152	27,62%
17	SVKDOLG_woe	705 067	27,43%	707 152	27,62%
18	RABKAPIT_woe	705 067	27,43%	707 152	27,62%

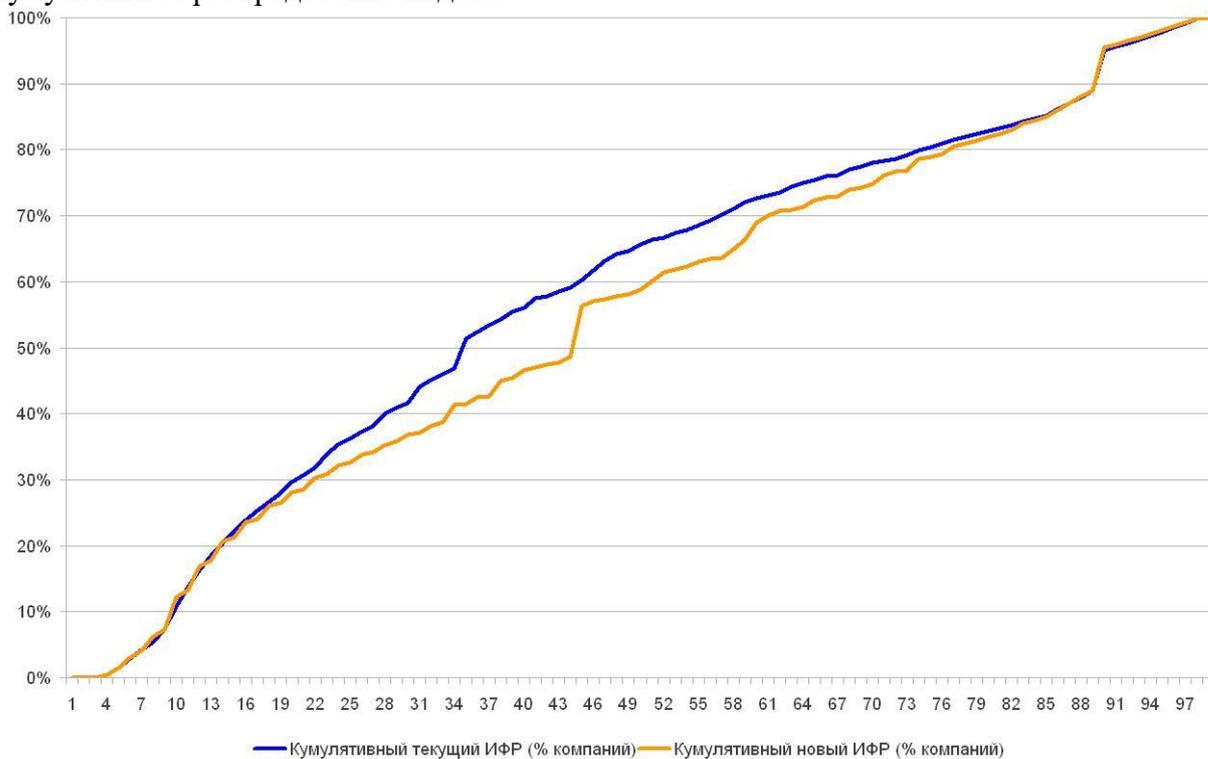
Устойчивость модели относительно версий

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

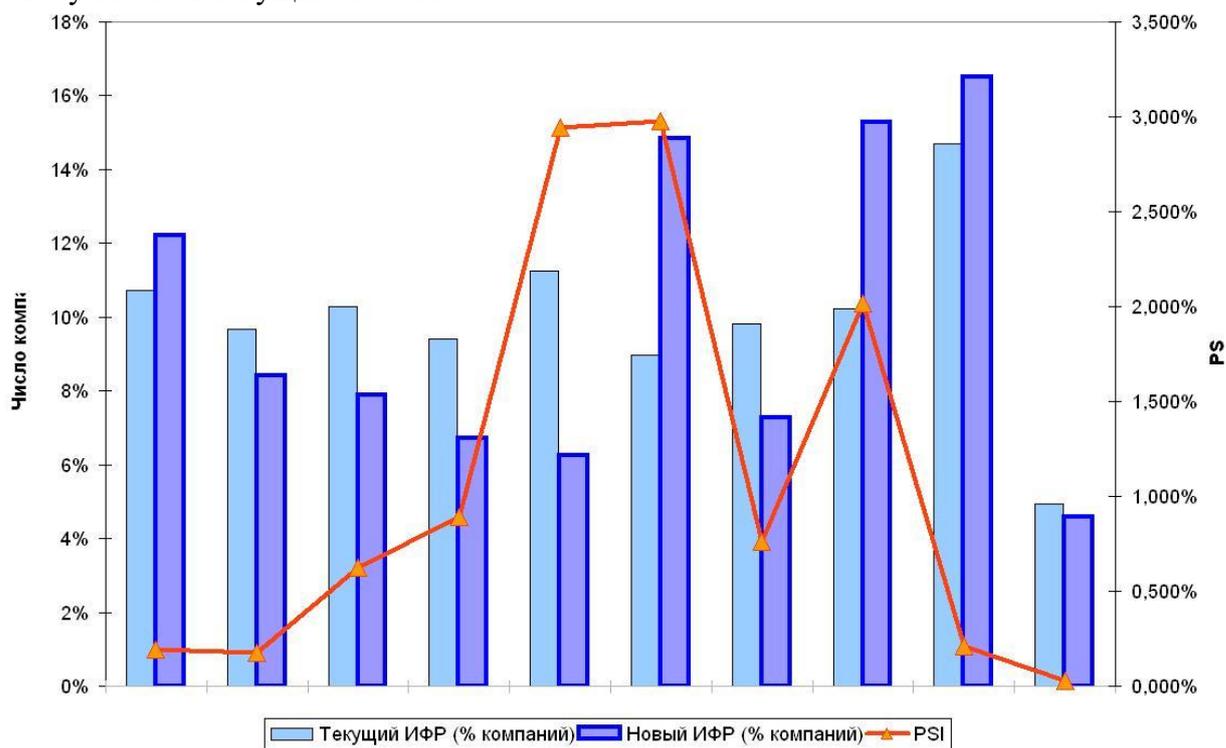
Распределение значений индекса:



Кумулятивное распределение индекса:

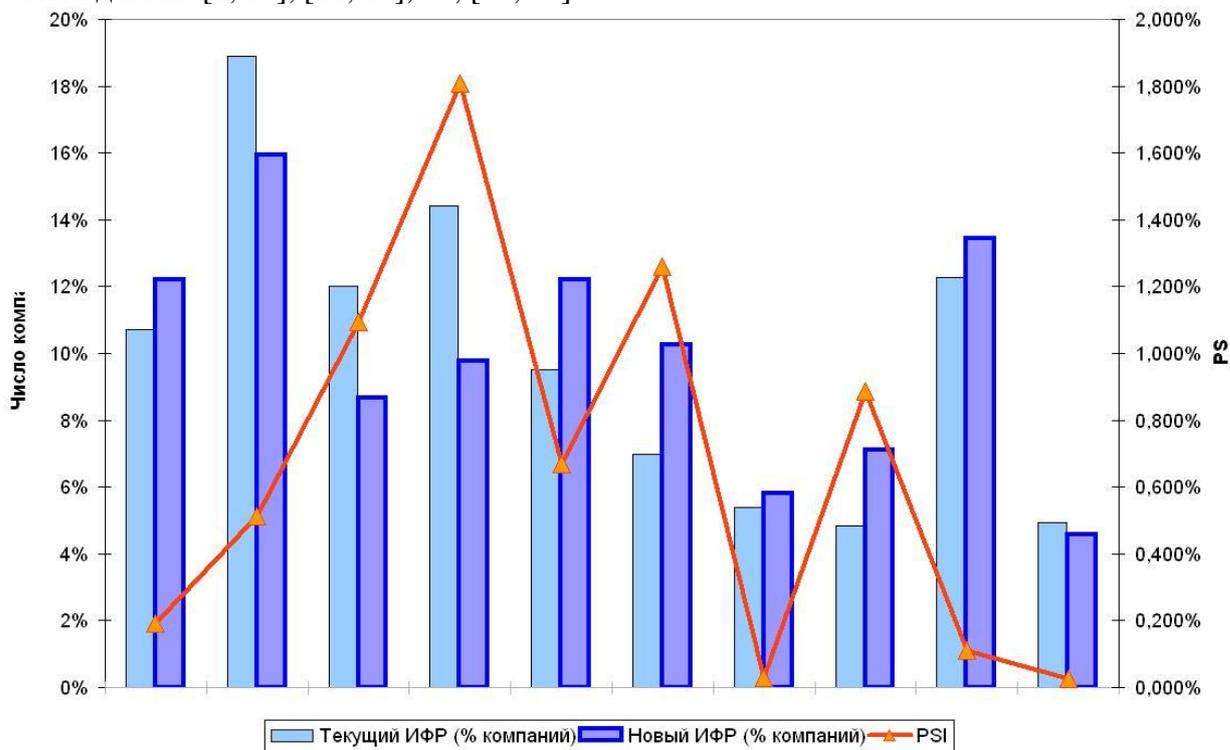


Распределение квантилей – значения индекса разбиты на 10 групп таким образом, чтобы в каждой из них находилось примерно равно число наблюдений для генеральной совокупности в текущий момент:



Итоговый PSI = 10,812%.

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



Итоговый PSI = 6,588%.