УДК: 330.42(045), 51-77(045)

ЕМКОСТНЫЙ МЕТОД АНАЛИЗА РЕДКИХ СОБЫТИЙ, ОЦЕНКА ПОГРЕШНОСТИ ВСЛЕДСТВИЕ КОНКУРЕНЦИИ ИЛИ ПОТЕРИ ДАННЫХ[1]

Кораблев Юрий Александрович, канд. экон. наук, доц.

Финансовый университет при Правительстве Российской Федерации (Финуниверситет), Ленинградский пр., 49, Москва, Россия, 125993; e-mail: yura-korablyov@yandex.ru

Цель: в статье кратко изложена идея емкостного метода анализа редких событий. Основное внимание уделено вопросу падения точности в результате потери части данных. *Обсуждение*: оценка погрешности проводилась с помощью компьютерного моделирования. Описана схема проведения экспериментов. Помимо этого, производилась оценка погрешности классическим методом, когда редкие события представлены временными рядами. В качестве оценки погрешности использовалась средняя относительная погрешность и средняя квадратичная погрешность. *Результаты*: исследование показало, что емкостный метод анализа редких событий обладает лучшей точностью по сравнению с классическим методом, независимо от того, какая часть данных теряется, 10% 30% или 50%. Однако в случае частых событий погрешность классического метода сравнивается. Делается вывод, что для анализа редких событий необходимо переходить на емкостный метод.

Ключевые слова: емкостный метод, конкуренция, потеря данных, моделирование, погрешность.

DOI: 10.17308/meps.2019.10/2222

1. Введение

Для работы с регулярными событиями существует большое количество проверенных и зарекомендовавших себя статистических методов: методы фильтрации и сглаживания [16], спектральный анализ [15], фазовый анализ [8, 240-247], методы аппроксимации и интерполяции [2], методы исследования временных рядов, факторный анализ [12, 13], адаптивные методы [8] и другие. Методов анализа редких событий очень мало по сравнению с методами анализа частых событий, классические методы обработки и анализа данных к редким событиям не подходят. Иногда применяют метод логистической регрессии [7], метод ближайшего соседа [9], метод Кростона

[1] Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №19-010-00154

[10, 14], метод бутстрэппинга (Виллемейна) [11, 17]. Иногда используются селективные методы [3]. Наиболее часто используют теорию случайных процессов [1], когда события представляются в виде Пуассоновского потока или обобщенного потока Эрганга, в более сложном варианте это может быть поток Пальма с ограниченным последействием. Однако перечисленные методы обладают низкой точностью, не способны предсказать сам момент наступления события (а лишь определить вероятность заданного количества на выбранном участке времени), поэтому разработка методов анализа редких событий является актуальной задачей.

В качестве подхода исследования редких событий предложена следующая методика: 1) разделяются источники событий; 2) делается предположение о характере процесса, протекающего в источнике, в результате которого образуются события; 3) по имеющимся данным происходит регрессия (восстановление) параметров процесса; 4) строятся модели и делается экстраполяция параметров процесса на будущее; 5) запускается сам процесс с установленными параметрами. Схема этого подхода анализа и прогнозирования редких событий представлена на рис. 1.

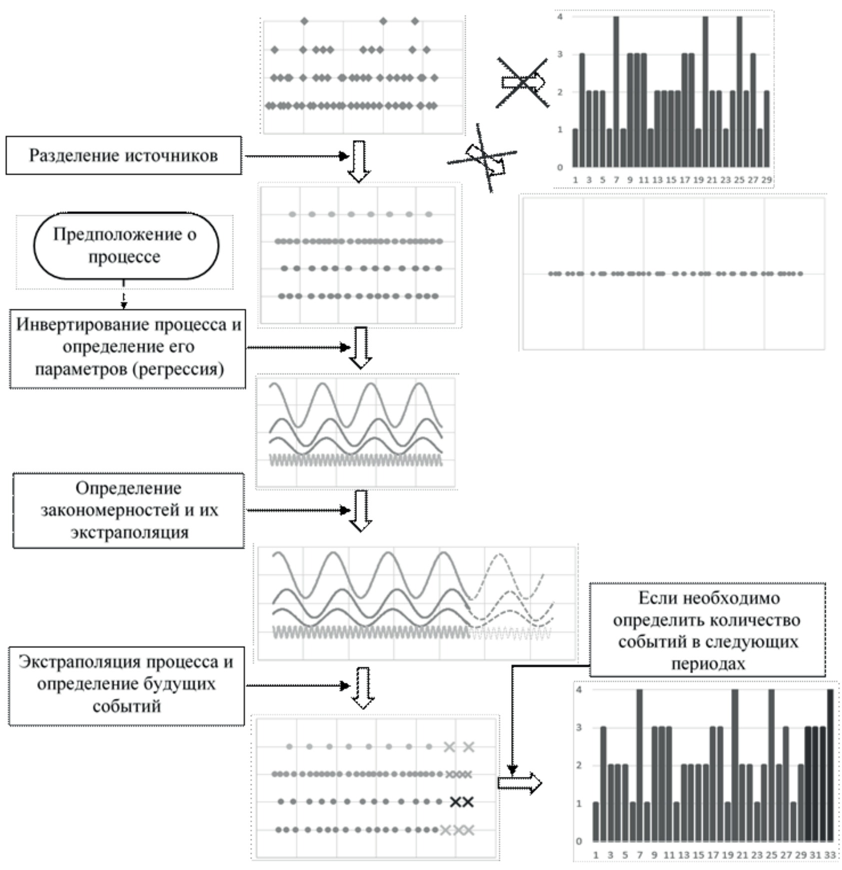

Рис. 1. Схема анализа и прогнозирования редких событий

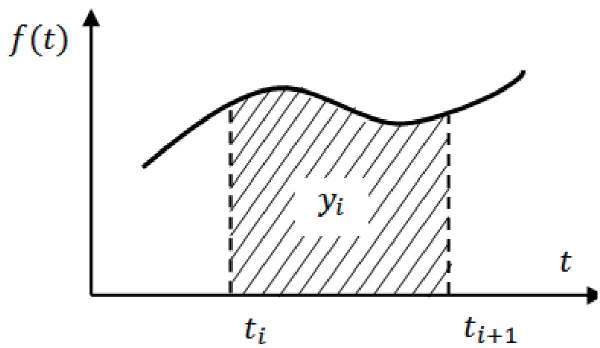

Рис. 2. Основное предположение

В экономике самым распространенным процессом является процесс потребления и расхода продукции, когда запас ведет себя как опустошающаяся емкость. В этом случае источники событий можно моделировать как емкости, предложенный метод анализа и прогнозирования редких событий я называю «ёмкостным методом». Параметром процесса является нестационарная функция скорости расхода запаса или накопления воздействия *f(t)*, подлежащая определению. Такой функцией может являться спрос от времени, индивидуальная скорость потребления продукции, интенсивность покупок у выбранного покупателя (источника, не путать со спросом или интенсивностью покупок у нас самих). Для восстановления этой функции используем основное предположение – величина совершенного события y_i есть интеграл от функции *f(t)* за время от момента совершения этого события t_i, до момента времени совершения следующего события t_{i+1}, рис. 2. Это предположение выполняется в случаях классического расхода или потребления продукции (как при моделировании системы управления запасами).

Задача определения (регрессии параметра процесса) *f(t)* представляет из себя оптимизационную задачу восстановления неизвестной функции, для которой известна последовательность интегралов за известные непересекающиеся периоды времени,

$$\sum_{i=1}^{n-1}\left(y_i-\int_{t_i}^{t_{i+1}} f(t)dt\right)^2+C\int_{t_1}^{t_n}\left(f''(t)\right)^2 dt \to \min,$$

где второе слагаемое – штраф за нелинейность; *C* – произвольная большая константа, влияющая на сглаживание; *n* – номер последнего события.

Мной было получено решение этой задачи, однако его изложение занимает много объема. Также можно предложить упрощенный способ, где достаточно ограничиться рассмотрением лишь средней скорости $y_i/(t_{i+1}-t_i)$, с которой происходят изменения запаса на каждом интервале, после чего сгладить полученную ступенчатую функцию любым известным методом для дальнейшего анализа. В этой статье будет использоваться упрощенный способ. Справедливость допущений не является темой этой статьи и здесь не рассматривается. Подробнее о емкостном методе можно посмотреть в работах [4-6].

2. Методы

Рассмотрим, как влияет потеря части данных на точность восстановления исходной зависимости (функции *f(t)*). Под потерей данных понимается случай, например, когда клиент решает совершить покупки не у того же самого распространителя, а у другого. На самом деле это приводит лишь к тому, что в выборке данных о продажах у распространителя будет отсутствовать часть данных. Эта картина очень легко моделируется. Например, при моделировании процесса расхода продукции можно с указанной вероятностью осуществлять покупку у одного из множества поставщиков. Тогда у каждого поставщика будет набор данных о покупках клиентов, который содержит пропущенные данные. В реальности это действительно происходит из-за конкуренции, когда клиенты меняют свое предпочтение между поставщиками. То есть можно говорить, что вследствие конкуренции происходит потеря части данных в истории покупок клиентов у распространителей. Если же брать лишь одного клиента, то моделировать потерю данных еще проще, достаточно из выборки с заданной вероятностью исключить каждую порцию данных о покупке.

Прежде чем говорить о точности, рассмотрим, например, следующую простую схему, рис. 3. Четыре конечных потребителя могут пополнять свои запасы у одного из двух вышестоящих распространителей (распространители, отмеченные номерами 1 и 2), которые в свое время могут пополнять свой запас продукции только у одного вышестоящего распространителя (распространитель, отмеченный номером 0). Пусть вероятность выбора конечными потребителями одного из распространителей будет равняться 50%. Также пусть исходная функция скорости расхода продукции у самого первого потребителя отличается частотой от скорости расхода продукции остальных трех (это нужно, чтобы лучше было видно, к чему приводит переключение потребителей).

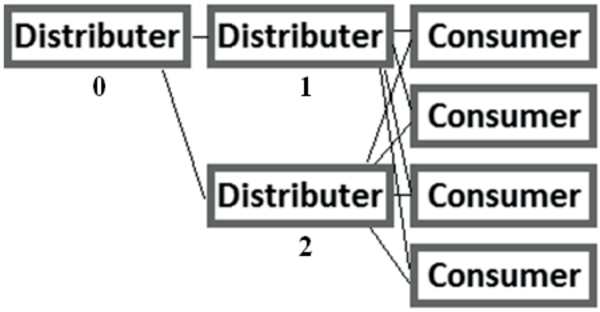

Рис. 3. Схема цепочки распространителей

На рис. 4 и 5 ступенчатой линией изображены восстановленные упрощенным способом функции скорости расхода продукции у распространителей 1 и 2. Гладкая черная линия показывает исходную (суммарную) скорость расхода продукции, у закрепленных за этим распространителей покупателей (суммарный расход продукции двумя потребителями).

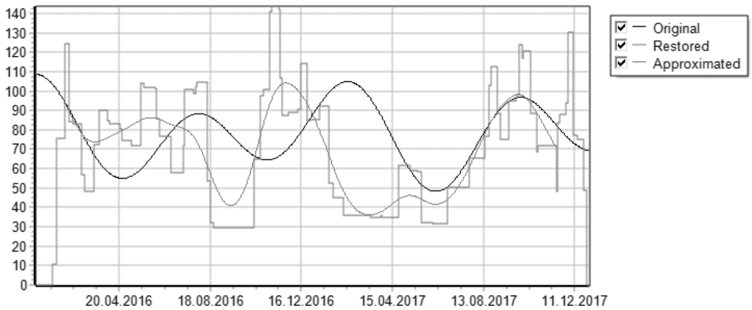

Рис. 4. Восстановленная скорость расхода продукции у распространителя 1

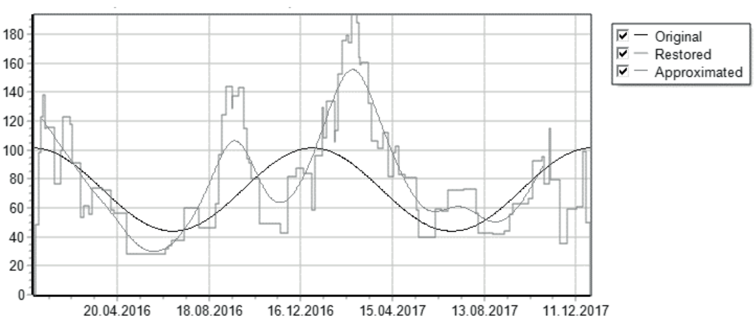

Рис. 5. Восстановленная скорость расхода продукции у распространителя 2

Визуально можно заметить, что там, где восстановленная ступенчатая функция ниже исходной на одном рисунке, на другом рисунке она выше и, наоборот, там, где восстановленная ступенчатая функция выше, на другом рисунке она ниже.

Теперь если взглянуть на рис. 6, где ступенчатой функцией показана восстановленная скорость расхода вышестоящего распространителя, отмеченного номером 0, можно увидеть, что восстановленная функция достаточно хорошо соответствует исходной скорости расхода продукции (сплошная линия, суммарная скорость расхода четырех потребителей, собственного расхода у распространителей нет). Неточность вызвана в основном удаленностью от конечного потребителя (появляется 1 уровень посредников).

Рис. 6. Восстановленная скорость расхода продукции у распространителя 0

В итоге можно сказать, что когда потребители переключаются между распространителями (продавцами) случайным образом, то в эти моменты восстановленная функция скорости расхода также переключается между этими распространителями. Когда такого переключения на более верхнем уровне не происходит, то есть когда все покупки нижестоящих распространителей совершаются только у одного вышестоящего (через несколько уровней посредников), то восстанавливается общая исходная скорость потребления, даже если на нижних уровнях это переключение происходило. Конечно же, такие результаты можно было бы заранее предсказать, сказав, что сколько вытягивается (реактивным методом как в системах управления запасами), то столько и должно восстанавливаться, но тем не менее подтвердить это наглядным экспериментом следовало.

Теперь перейдем к сравнению точности. Во-первых, надо определиться, с чем сравнивать и по какому критерию. Сравнивать будем с классическим подходом, когда из данных строится временной ряд, например, продаж по месяцам. Однако так наша задача определить точность восстановления исходной функции (скорости расхода запаса), то будем считать скорость расхода как отношение объема продаж в этом месяце к количеству дней в этом месяце. То есть временной ряд превратим в ступенчатую функцию скорости расхода продукции в каждый месяц. Сглаживание производить не будем. Так как классический подход будет выглядеть как ступенчатая функция, то и восстановленную упрощенным способом, емкостным методом функцию также не будем сглаживать, функция будет ступенчатой.

Так как исходная функция не стационарна, то будем определять среднюю относительную ошибку (модуль разницы восстановленной и истинной функции, деленный на истинное значение). С другой стороны, в статистике наиболее популярным является определение средней квадратичной ошибки, дополнительно будем измерять среднюю квадратичную относительную ошибку.

Для получения результатов создадим модель, в которой в каждом отдельном прогоне случайным образом задается исходная функция расхода продукции (средний уровень, амплитуда, частота и фаза задаются случайными числами из некоторого промежутка), случайным образом задается начальный запас, максимальный запас. Затем запуская классический процесс расхода продукции (как в управлении запасами). Моделирование проводить будем за 2 года. Будем изменять вероятность P того, что клиент обратится к другому поставщику и данные потеряются (информация о покупке пропадет). Разыгрывание вероятности также происходит случайным образом, то есть при вероятности 10% потери данных из выборки в 20 наблюдений может потеряться как 2, так и 1, 0 или же 3 и 4 наблюдения (количество потерянных наблюдений соответствует биномиальному распределению). В то же время есть вероятность того, что потери будут идти подряд (нет строгого чередования).

Пример моделирования показан на рис. 7. В этом примере вероятность потери данных была 10%. Из 22 наблюдений было потеряно лишь одно, потерянное наблюдение приходилось на 9-й месяц. Пунктирная ступенчатая линия показывает восстановленную функцию с помощью емкостного метода. Сплошная ступенчатая линия показывает временной ряд (площадь ступеньки есть продажи за месяц), преобразованный в скорость расхода продукции. Гладкая сплошная линия показывает истинную скорость расхода продукции, которую требовалось восстановить из данных о продажах. В этом прогоне средняя относительная погрешность восстановленной функции составила 11,9% для емкостного метода и 58,4% для классического, средняя квадратичная относительная погрешность составила 18,7% и 70,1% соответственно.

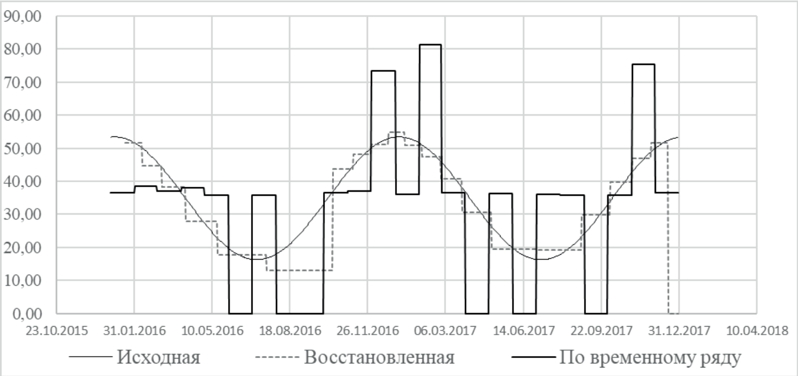

Рис. 7. Пример моделирования для оценки погрешности

3. Результаты

В зависимости от многих факторов результаты моделирования могут быть разными. В табл. 1 приведены результаты 20 прогонов для вероятности потери данных 10%.

Таблица 1

Относительная погрешность при 10% вероятности потери данных

№	Средняя относительная ошибка в %		Разница	Средняя квадратичная относительная ошибка в %		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	15,1538	35,6751	20,5213	25,7686	40,5931	14,8245
2	19,2385	42,6782	23,4397	31,4277	55,4565	24,0287
3	10,4954	23,9023	13,4069	20,6358	28,7282	08,0924
4	20,1894	41,3341	21,1447	32,1758	53,1899	21,0141
5	25,6441	55,8185	30,1744	34,7822	87,6615	52,8793
6	16,5491	36,9385	20,3894	27,5394	49,4287	21,8893
7	11,0051	22,9302	11,9251	19,8246	30,479	10,6544
8	20,6452	70,037	49,3918	29,0427	101,1585	72,1158
9	10,1343	24,5582	14,424	19,2464	31,128	11,8816
10	14,3658	38,2592	23,8934	26,9225	48,9228	22,0003
11	13,921	48,0131	34,0921	22,801	63,1846	40,3836
12	6,8549	46,397	39,5421	9,1229	60,8571	51,7342

Окончание табл. 1

№	Средняя относительная ошибка в %		Разница	Средняя квадратичная относительная ошибка в %		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
13	20,5882	28,2138	07,6256	32,345	38,6496	06,3046
14	6,7678	26,0652	19,2974	12,4631	33,1063	20,6432
15	13,3678	19,7823	06,4146	24,2151	27,0365	02,8215
16	10,7269	24,8431	14,1162	20,0973	34,3362	14,2389
17	15,9477	38,1086	22,1609	26,0582	49,6658	23,6076
18	15,7876	46,3486	30,561	25,6549	58,6784	33,0235
19	14,9275	58,5897	43,6622	23,3633	71,5303	48,167
20	19,536	69,6921	50,1562	25,8494	94,0599	68,2105

Классический метод полностью проигрывает в каждом эксперименте. Относительная погрешность емкостного метода в среднем на 24,8% меньше (на 28,4 для средней квадратичной относительной погрешности).

Проведем аналогичное моделирование с вероятностью потери данных 30%. Результаты представлены в табл. 2.

Таблица 2

Относительная погрешность при 30% вероятности потери данных

№	Средняя относительная ошибка в %		Разница	Средняя квадратичная относительная ошибка в %		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	40,661	49,478	8,817	45,467	60,349	14,882
2	38,775	39,536	0,761	48,271	49,835	1,564
3	29,129	59,977	30,848	36,811	72,629	35,818
4	26,551	35,727	9,176	39,613	46,353	6,740
5	27,419	32,571	5,152	38,257	46,821	8,564
6	24,549	35,926	11,378	36,321	43,810	7,489
7	26,360	30,149	3,790	37,601	37,165	-0,435
8	33,776	39,033	5,257	44,006	52,257	8,252
9	33,975	37,804	3,829	44,284	44,110	-0,174
10	38,246	54,343	16,096	45,526	64,534	19,008
11	27,181	50,703	23,522	37,889	61,949	24,059
12	16,978	54,400	37,422	26,519	66,288	39,769
13	30,821	41,901	11,081	40,159	54,406	14,248
14	47,786	60,916	13,130	58,851	72,022	13,171
15	37,405	49,773	12,368	47,676	60,656	12,980
16	32,871	46,228	13,357	42,424	57,839	15,415
17	36,227	37,889	1,662	45,674	45,668	-0,006
18	38,379	72,399	34,019	46,106	86,715	40,609
19	13,814	67,873	54,059	19,269	84,973	65,704
20	31,854	36,204	4,351	39,892	42,649	2,757

При вероятности потери данных 30% ошибка восстановления исходной функции уже большая, тем не менее в каждом эксперименте средняя относительная ошибка емкостного метода меньше (в среднем на 15%), средняя квадратичная относительная в основном меньше (в среднем на 16,5%), правда, кое-где ошибка сравнялась.

Наконец проведем моделирование, в котором вероятность потери данных будет 50%. Результаты представлены в табл. 3.

Таблица 3

Относительная погрешность при 50% вероятности потери данных

№	Средняя относительная ошибка в %		Разница	Средняя квадратичная относительная ошибка в %		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	47,746	56,215	8,468	27,933	44,038	16,104
2	35,898	71,731	35,833	44,054	81,189	37,135
3	53,394	74,495	21,102	60,303	82,048	21,745
4	35,768	68,061	32,293	43,550	77,317	33,767
5	52,566	84,793	32,227	62,241	93,685	31,444
6	51,347	63,975	12,628	57,779	75,714	17,935
7	44,194	61,661	17,467	51,086	72,101	21,015
8	61,588	62,980	1,392	65,735	74,196	8,461
9	51,375	70,069	18,695	56,987	81,231	24,244
10	46,771	73,436	26,664	56,859	79,695	22,836
11	47,270	66,255	18,985	55,215	72,996	17,781
12	39,825	45,107	5,282	46,423	57,808	11,384
13	45,203	66,544	21,341	54,167	73,718	19,551
14	59,472	77,924	18,452	64,622	86,228	21,606
15	54,556	79,175	24,620	61,530	86,874	25,344
16	48,729	70,110	21,381	55,809	77,713	21,904
17	52,266	62,035	9,769	62,942	74,023	11,081
18	53,202	62,206	9,004	61,232	72,977	11,746
19	34,200	78,717	44,517	41,817	91,231	49,414
20	57,245	84,039	26,794	64,176	94,453	30,277

Погрешность стала совсем огромной, однако опять в каждом эксперименте средняя относительная ошибка емкостного метода оказалась меньше (в среднем на 20,3%), средняя квадратичная относительная ошибка также меньше (в среднем на 22,7%), правда, кое-где ошибка сравнялась.

Во всех трех сериях экспериментов наглядно видно, что емкостный метод обладает меньшей погрешностью при потерях данных, следовательно, при анализе редких событий (продаж) по возможности надо всегда использовать его, а не строить временные ряды с нулями продаж за выбранный период.

Однако, что, если мы имеем дело не с редкими продажами, а с частыми, например, когда каждый покупатель делает много покупок за месяц. Как в этом случае будет изменяться точность при потере части данных. Временной ряд не будет содержать нулей, каждое пропущенное наблюдение лишь уменьшает величину расхода за месяц. Пример моделирования для вероятности потери данных 10% изображен на рис. 8. Результаты моделирования потери данных с вероятностью 10% при частых продажах показаны в табл. 4.

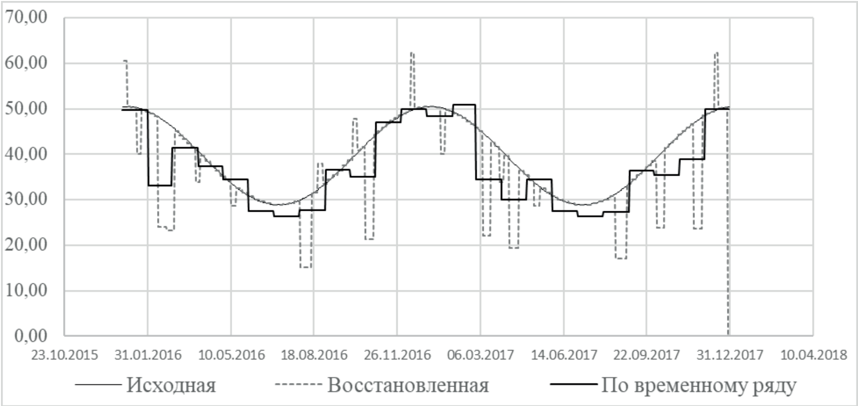

Рис. 8. Пример моделирования для оценки погрешности при частых продажах

Таблица 4

Относительная погрешность при 10% вероятности потери данных при частых продажах

№	Средняя относительная ошибка в %		Разница	Средняя квадратичная относительная ошибка в %		Разница
	Емкостный метод	Классический метод		Емкостный метод	Классический метод	
1	11,252	11,424	0,172	21,540	16,222	-5,318
2	14,576	16,303	1,727	25,636	20,293	-5,343
3	9,628	11,772	2,144	18,070	16,242	-1,828
4	13,941	22,024	8,083	20,583	29,383	8,801
5	8,484	9,130	0,646	18,776	11,951	-6,825
6	11,813	12,338	0,526	22,886	16,999	-5,887
7	16,154	19,831	3,676	24,162	26,511	2,350
8	11,650	14,152	2,502	22,791	18,475	-4,316
9	13,174	14,780	1,605	25,119	21,557	-3,562
10	12,732	13,664	0,932	23,989	17,287	-6,702

В каждом эксперименте средняя относительная погрешность емкостного метода по-прежнему меньше. Однако средняя квадратичная относительная погрешность – в большинстве случаев, наоборот, больше. Это связано с тем, что из-за потери данных при частых продажах емкостный метод показывает занижение на узком интервале времени, которое достаточно значительно по величине. Возведение в квадрат этой относительной разницы вносит большой вклад в оценку погрешности, даже при условии, что эти спады кратковременны. Стоит заметить, что любое сглаживание способно решить эту проблему.

4. Заключение

В результате исследования было показано, что емкостный метод анализа редких событий в случае потери части данных теряет меньше в точности восстановления исходной зависимости, чем классический метод, базирующийся на построении временных рядов. В каждом проведенном эксперименте средняя относительная погрешность емкостного метода была меньше, чем у классического. В случае анализа частых событий, когда временной ряд не содержит нулевых значений, я по-прежнему рекомендую использовать емкостный метод, так как он способен дать много информации о протекающих в источниках процессах. В следующих работах я постараюсь математически обосновать результаты данного исследования, чтобы на языке логики и математики выявленные закономерности подтвердились.

Список источников

1. Вентцель Е.С., Овчаров Л.А. *Теория случайных процессов и ее инженерные приложения*: учеб. пособие для втузов. 2-е изд., стер. Москва, Высш. шк., 2000.

2. Голубинский А.Н. Методы аппроксимации экспериментальных данных и построения моделей // *Вестник Воронежского института МВД России*, 2007, no. (2), с. 138-143.

3. Иванько Р.С. *Краткосрочное прогнозирование нестационарного спроса в оптовой торговле*: дис. … канд. эконом. наук. Москва, 2005.

4. Кораблев Ю.А. Емкостный метод анализа редких событий в торговле различными товарами // *Бизнес. Образование. Право. Вестник Волгоградского института бизнеса*, 2019, no. 3(48), с. 121-131.

5. Кораблев Ю.А. Емкостный метод определения функции скорости потребления // *«Экономика и менеджмент систем управления»*. Воронеж, Научная книга, 2015, no. 15(1.1), с. 140-150.

6. Кораблев Ю.А. Погрешность емкостного метода анализа редких событий, удаленность от конечного потребителя // *Известия Кабардино-Балкарского научного центра РАН*. Нальчик, Известия КБНЦ РАН, 2019, no. 3 (89), с. 48-77.

7. Корректировка числа редких событий в логистической регрессии. Доступно: http://www.statmethods.ru/stati/178-korrektirovka-chisla-redkikh-sobytij-v-modeli-logisticheskoj-regressii.html (дата обращения: 03.03.2019).

8. Лукашин Ю.П. *Адаптивные методы краткосрочного прогнозирования временных рядов*. Москва, Финансы и статистика, 2003.

9. Altman N.S. An introduction to kernel and nearest-neighbor nonparametric regression // *The American Statistician*, 1992, no. 46(3), pp. 175-185.

10. Croston J.D. Forecasting and stock control for intermittent demands // *Operational Research Quarterly* (1970-1977), 1972, no. 23(3), pp. 289-303.

11. Efron B. and R.J. Tibshirani. *An introduction of the Bootstrap*. New York, Chapman & Hall, 1993.

12. Jackson J.E. Principal components and factor analysis: Part I-principal components // *Journal of Quality Technology*, 1980, no. (12), pp. 201-213.

13. Jackson J.E. Principal components and factor analysis: Part II-additional topics related to principal components // *Journal of Quality Technology*, 1981, no. (13), pp. 46-58.

14. Johnston F.R., Boylan J.E. Forecasting intermittent demand: a comparative evaluation of Croston's method. Comment // *International journal of forecastin*g, 1996, no. 12(2), pp. 297-298.

15. Lambert Koopmans. *The Spectral Analysis of Time Series.* 1st Edition. University of New Mexico. Academic Press. 1995. Доступно: https://www.elsevier.

com/books/the-spectral-analysis-of-time-series/koopmans/978-0-12-419251-5/ (дата обращения: 03.03.2019).

16. Saerkkae S. *Bayesian Filtering and Smoothing.* Cambridge University Press. 2013. Доступно: http://www.cambridge.org/9781107030657/ (дата обращения: 03.03.2019).

17. Willemain T.R., Park D.S., Kim Y.B., Shin K.I. *Simulation output analysis using the threshold bootstrap*, 2001, no. 134(1) pp. 17-28.

CAPACITY METHOD FOR ANALYZING RARE EVENTS, ERROR ESTIMATION DUE TO COMPETITION OR DATA LOSS

Korablev Juri Alexandrovich, Cand. Sc. (Econ.), Assoc. Prof.

Financial University under the Government of the Russian Federation, Leningradskii pr., 49, Moscow, Russia, 125993; e-mail: yura-korablyov@yandex.ru

Purpose: the article briefly describes the idea of a capacity method for analyzing rare events. The article focusses on the drop in accuracy as a result of the data loss. *Discussion*: error estimation was carried out using computer simulation. The authors describe experimental model, In addition, estimate the error by classical methods, when rare events are represented by time series with zero values. As an estimate of the error, the authors use the average relative error and the mean square relative error. *Results*: the study shows that the capacity method of rare events analysis has better accuracy compared to the classical method, regardless of how much of the data is lost, 10% 30% or 50%. However, in the case of frequent events, the error of the classical method is almost the same. The authors conclude that for the rare events analysis it is necessary to switch to the capacity method.

Keywords: capacity method, competition, data loss, modeling, error.

References

1. Wentzel E.S., Ovcharov L.A. *Teoriya sluchainyh protsessov i ee inzhenernye prilozheniya*: ucheb. posobie dlya vtuzov [The theory of random processes and its engineering applications. – Training. manual for technical colleges]. 2nd ed., Sr. Moscow, Higher. school, 2000. (In Russ.)

2. Golubinsky A.N. Metody approksimatsii eksperimentalnyh dannyh i postroeniya modelei [Experimental data approximation and model building methods]. *The bulletin of Voronezh Institute of the Ministry of Internal Affairs of Russia*, 2007, no. (2), pp. 138-143. (In Russ.)

3. Ivanko R.S. *Kratkosrochnoe prognozirovanie nestatsionarnogo sprosa v optovoi torgovle*: dis kand. ekonom. nauk [Short-term forecasting of non-stationary demand for wholesale: dis. Cand. economy sciences]. Moscow, 2005. (In Russ.)

4. Korablev Yu.A., Golovanova P.S., Kostritsa T.A. Emkostnyi metod analiza redkih sobytii v torgovle razlichnymi tovarami [Capacity method of analyzing rare events in the trade of various goods]. *Business. Education. Law*, 2019, no. 3, pp. 121-131. (In Russ.)

5. Korablev Yu.A. Emkostnyi metod opredeleniya funktsii skorosti potrebleniya [Capacity method determination consumption rate function]. *«Ekonomika i menedzhment sistem upravleniya»*. Voronezh, Nauchnaya kniga, 2015, no. 15(1.1), pp. 140-150. (In Russ.)

6. Korablev Yu.A. Pogreshnost emkostnogo metoda analiza redkih sobytii, udalennost ot konechnogo potrebitelya [Error of the capacity method of rare events analysis, remoteness from the end user]. *The News of KBSC of RAS*, 2019, no. 3 (89), pp. 48-77. (In Russ.)

7. [Rare events number correction in the logistic regression]. (In Russ.) Available at:www.statmethods.ru/stati/178-korrektirovka-chisla-redkikh-sobytij-v-modeli-

logisticheskoj-regressii.html (accessed: 03.03.2019).

8. Lukashin Yu.P. *Adaptivnye metody kratkosrochnogo prognozirovaniya vremennyh ryadov* [Adaptive methods for short-term time series forecasting]. Moscow, Finansy i statistika, 2003. (In Russ.)

9. Altman N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 1992, no. 46(3), pp. 175-185.

10. Croston J.D. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* (1970-1977), 1972, no. 23(3), pp. 289-303.

11. Efron B. and R.J. *Tibshirani, An introduction of the Bootstrap*. New York, Chapman & Hall, 1993.

12. Jackson J.E. Principal components and factor analysis: Part I-principal components. *Journal of Quality Technology*, 1980, no. (12), 201-213.

13. Jackson J.E. Principal components and factor analysis: Part II-additional topics related to principal components. *Journal of Quality Technology*, 1981, no. (13), pp. 46-58.

14. Johnston F.R., Boylan J.E. Forecasting intermittent demand: a comparative evaluation of Croston's method. Comment. *International journal of forecasting*, 1996, no. 12(2), pp. 297-298.

15. Lambert Koopmans. *The Spectral Analysis of Time Series.* 1st Edition. University of New Mexico. Academic Press, 1995. Available at: https://www.elsevier.com/books/the-spectral-analysis-of-time-series/koopmans/978-0-12-419251-5/ (accessed: 03.03.2019).

16. Saerkkae S. *Bayesian Filtering and Smoothing.* Cambridge University Press. 2013. Available at: http://www.cambridge.org/9781107030657/ (accessed: 03.03.2019).

17. Willemain T.R., Park D.S., Kim Y.B., Shin K.I. *Simulation output analysis using the threshold bootstrap*, 2001, no. 134(1), pp. 17-28.