

УДК: 330.42(045), 51-77(045)

JEL: C1, C15, C4, C5, C53

DOI: 10.35330/1991-6639-2020-5-97-49-67

## ИССЛЕДОВАНИЕ ТОЧНОСТИ ЕМКОСТНОГО МЕТОДА АНАЛИЗА РЕДКИХ СОБЫТИЙ ОТ НЕОПРЕДЕЛЕННОСТИ ВНУТРИ ПРОЦЕССА ОБРАЗОВАНИЯ СОБЫТИЙ\*

(Часть 1)

Ю.А. КОРАБЛЕВ

Финансовый университет при Правительстве Российской Федерации (Финуниверситет)  
125993, г. Москва, Ленинградский проспект, 49  
E-mail: academy@fa.ru

*В исследовании определяется влияние неопределенности внутри процесса потребления образования редких событий на точность восстановления исходной функции потребления по данным редких событий с помощью емкостного метода. Восстановление функции происходит по последовательности интегралов, наблюдаемых с погрешностью, с использованием кубического интегрального сглаживающего базисного сплайна. В данной части исследования в первом эксперименте проверяется влияние погрешности в самих наблюдениях на точность восстановления функции. Во втором эксперименте определяется влияние случайного разброса во время проверки уровня запасов, а в третьем эксперименте определяется влияние разброса в датах самих событий на фактическую погрешность наблюдений и на погрешность восстановления исходной функции.*

*Эксперименты показали, что неопределенность внутри самого процесса образования событий влияет лишь на погрешность в наблюдениях, которая в свою очередь влияет линейно на погрешность восстанавливаемой функции. Построена модель этой линейной зависимости. В то же время большая погрешность наблюдений приводит к появлению шума на восстанавливаемой функции. Предложен механизм удаления этого шума с помощью повышения коэффициента сглаживания в процессе восстановления функции по интегралам.*

**Ключевые слова:** редкие события, емкостный метод, процесс потребления, скорость потребления, восстановление, регрессия, интегральный сплайн, погрешность, точность.

### 1. ВВЕДЕНИЕ

Ранее для анализа редких событий был разработан емкостный метод [1, 2], идея которого заключается в том, чтобы анализировать события с точки зрения процессов, протекающих внутри источников событий и приводящих к образованию этих событий. Из данных редких событий получается восстановить параметры таких процессов, после чего можно проводить дальнейшие исследования зависимостей и экстраполяцию, давать прогноз появления будущих событий. Для процесса потребления, который является самым распространенным процессом образования событий в экономике, по данным о совершенных покупках можно восстановить скорость, с которой расходовался запас у совершающих эти покупки покупателей. Запас продукции в этом случае моделируется как опустошающая/пополняющаяся емкость, объемы покупок – это величины, на которые пополняется емкость, скорость потребления – это скорость опустошения емкости. Интерес представляет эта скорость опустошения емкости, которая может иметь нестационарный характер. Используя предположение, что наблюдаемый объем пополнения емкости есть интеграл от неизвестной искомой функции от момента пополнения до момента следующего пополнения, мы можем восстановить эту

---

\* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №19-010-00154

функцию. Для этого необходимо решать задачу восстановления функции по интегралам, где интегралами выступают объемы покупок. Для решения этой задачи используется кубический интегральный сглаживающий базисный сплайн [3], с помощью которого можно построить функцию по известной последовательности интегралов от функции, которые могут наблюдаться с погрешностью  $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$ .

Погрешность наблюдаемых интегралов сказывается на точности восстанавливаемой из этих интегралов функции. Читателю может быть интересно, откуда берется эта погрешность в наблюдаемых интегралах при функционировании процесса потребления. Погрешность возникает вследствие многих причин. Это и использование страховых запасов, когда часть продукции берется из этих страховых запасов, а часть покупки откладывается для восполнения страховых запасов [4]. Также поведение потребителя не всегда идеально, он может не своевременно пополнять запасы, а чуть позже или чуть раньше. В результате всего этого наблюдаемый объем покупки не соответствует реально расходуемому объему продукции, что и приводит к погрешности. Причем надо заметить, что в отличие от погрешности, которая присутствует в таких дисциплинах, как эконометрика, где погрешность показывает отклонение значений, полученных по некоторой модели, от наблюдаемых данных (то есть наблюдения происходят без погрешности, а эта модель обладает погрешностью), у нас именно сами наблюдения происходят с погрешностью. Однако на решении это отличие не сказывается, восстановление функции по интегралам происходит с помощью минимизации суммы квадратов разницы между наблюдениями и рассчитанными интегралами от восстанавливаемой функции.

В данном исследовании мы попытаемся определить три аспекта. Первый: влияние самой погрешности наблюдаемых интегралов на точность восстанавливаемой функции. Далее мы попытаемся определить влияние неопределенности внутри процесса образования событий на погрешность наблюдаемых интегралов, которая приводит к погрешности восстановления исходной функции из этих интегралов. Причем в процессе образования событий (которым будет процесс потребления) такой неопределенностью у нас будет неопределенность при проверке уровня запасов на то, опустились ли они ниже критического уровня запасов или нет. Третий аспект, который мы проверим, будет неопределенность в наблюдаемых датах событий, что, по сути, сводится к той же погрешности наблюдаемых интегралов. Исследование будем проводить с помощью имитационного моделирования, когда мы можем априори задать исходную функцию потребления, которую затем будем восстанавливать из последовательности наблюдаемых интегралов.

#### МЕТОД ВОССТАНОВЛЕНИЯ ФУНКЦИИ ПО ИНТЕГРАЛАМ

Прежде чем непосредственно переходить к результатам исследования, необходимо несколько слов сказать о самом методе восстановления функции по интегралам [3]. Для этого решается следующая оптимизационная задача:

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

где  $g(t) = \sum_{k=1}^m \delta_k \beta_k(t)$  – кубический базисный сплайн, представленный комбинацией  $m$  базисных функций  $\beta_k(t)$ , которые являются полиномами третьей степени, определенными между соседними узлами сплайна  $s_1 \leq \dots \leq s_m$ , но выражаются не через 4 коэффициента, а через значения самого сплайна  $g(s_k)$  и его второй производной  $g''(s_k)$  в узлах сплайна (см. *value-second derivative representation* в [5, с. 12]),  $\delta_k$  – коэффициент, принимающий значение 0 или 1 в зависимости от того, какая функция соответствует текущему моменту времени  $t$ ; вторая часть суммы есть штраф на нелинейность,  $\alpha$  – коэффициент, влияющий на сглаживание.

Особенностью данного сплайна (как и любого другого натурального) является то, что на концах всего интервала вторая производная обращается в ноль  $g''(s_1) = g''(s_m) = 0$ , что

приводит к тому, что функция старается превратиться в прямую линию при подходе к концам области определения. Такое происходит из-за того, что у нас нет данных раньше, чем самое первое наблюдение, и позже, чем самое последнее наблюдение. Это приводит к тому, что на самом первом и последнем участке погрешность будет как правило самая большая.

Коэффициент сглаживания  $\alpha$  также имеет очень сильное влияние на восстанавливаемую функцию. При  $\alpha \rightarrow 0$  сглаживания совсем не происходит, при  $\alpha \rightarrow \infty$  функция превращается в прямую линию. Заметим, что при  $\alpha = 0$  задача восстановления функции становится «некорректной» в том смысле, что она может иметь бесконечное множество решений, можно произвольным образом провести функцию, чтобы площадь под ней совпала с известными интегралами (параметр  $\alpha$  в относительно старой отечественной литературе 70-80-х годов называют «альфа-регуляризацией» Тихонова для решения некорректных задач [6], однако в иностранных статьях, посвященных сглаживающим сплайнам, об этом не упоминается, параметр  $\alpha$  называют просто параметром сглаживания). Параметр  $\alpha$  надо выбирать таким образом, чтобы функция сглаживала отклонения, вызываемые погрешностью наблюдения, но не сглаживалась сама исходная функция. Добиться этого непросто. Иногда используют автоматический механизм выбора сглаживающего параметра на основе Кросс-Валидации, но для нашей задачи восстановления функции по интегралам с помощью сплайнов такого механизма еще нет, кроме того, такой автоматический подход не всегда дает хороший результат. Параметр  $\alpha$ , по сути, ограничивает максимальное ускорение функции, не позволяя слишком резко возрастать и убывать. Причем штраф на нелинейность может никак не проявляться до тех пор, пока параметр  $\alpha$  не примет большие значения. Такое может произойти из-за того, что отклонение функции от своего положения вызовет сразу значительное увеличение суммы квадратов отклонений, компенсировать которое необходимо большим штрафом на нелинейность (такое наблюдается в слабо зашумленных данных). Одним из подходов по выбору параметра сглаживания можно предложить визуальное сравнение получаемой функции со ступенчатым графиком средних значений. Если на графике восстановленная функция сглаживает случайные отклонения, но не сглаживает саму функцию (тут появляется субъективная оценка, так как мы как правило не знаем, как выглядит исходная функция), значит, параметр подобран верно.

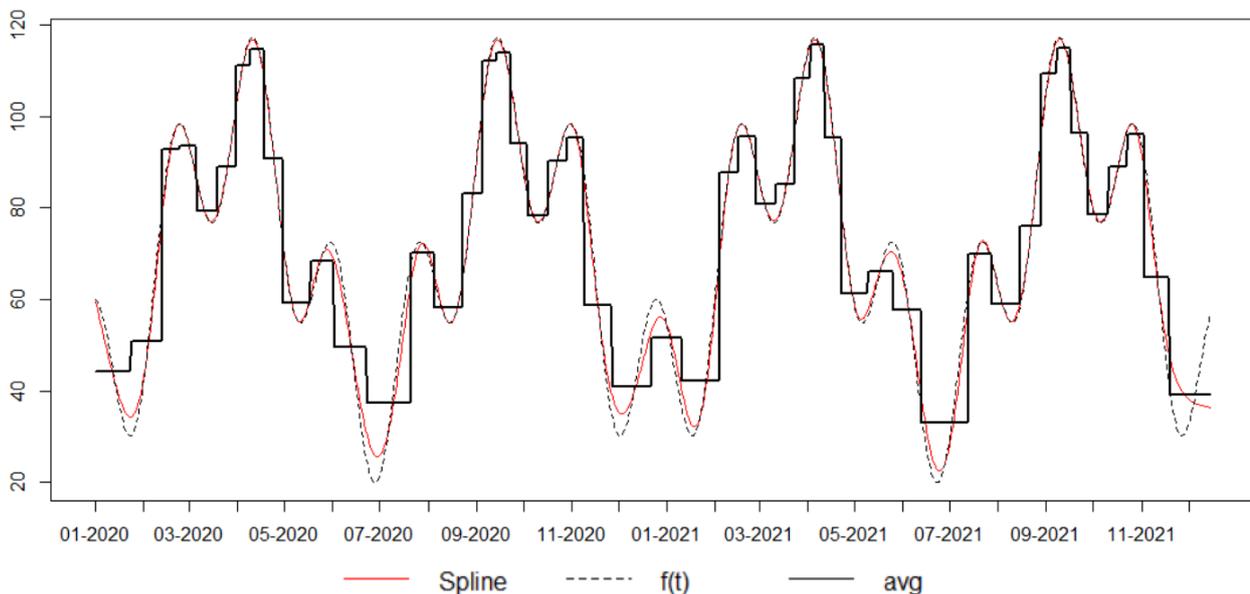
#### МЕТОД ИССЛЕДОВАНИЯ ТОЧНОСТИ

Производить исследование будем с помощью имитационного моделирования, когда мы можем заранее задать исходную функцию. Моделировать мы будем процесс потребления продукции так же, как при моделировании систем управления запасами. Зададим скорость расхода продукции некоторого потребителя в виде определенной функции, зададим максимальный уровень запасов, критический уровень, после которого происходит пополнение запасов до максимума, пусть пополнение запаса происходит мгновенно (игнорируем время доставки заказа). В результате моделирования мы получаем моменты времени  $t_i$  и объемы пополнения запаса  $u_i$ . Помимо этого, также получим точные значения интегралов от исходной функции за время между моментами пополнения запаса  $I_i = \int_{t_i}^{t_{i+1}} f(t) dt$ . Эти интегралы не будут совпадать с объемами пополнения запаса, так как часть от этого объема будет компенсировать запасы, которые опустились ниже критического уровня. По данным моделирования будем восстанавливать исходную функцию  $\hat{f}(t)$  из наблюдаемых объемов пополнения запаса (параметр сглаживания  $\alpha$  определим визуально один раз и для разных экспериментов будем использовать одно и то же значение). Для определения точности восстановления исходной функции  $f(t)$  будем находить среднюю относительную погрешность восстановления в процентах:

$$R = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{f}(t) - f(t)|}{f(t)} \times 100\%.$$

Вместе с тем будем определять  $\sigma_{\text{факт}}^{\text{набл}}$  фактическое СКО относительного отклонения наблюдаемых объемов пополнения запасов от рассчитанных аналитически интегралов  $\frac{y_i - I_i}{I_i}$  и попробуем показать связь между относительной погрешностью восстановления функции и этим разбросом в наблюдениях.

Обратим внимание, что если мы восстанавливаем функцию по истинным интегралам  $I_i$  от функции, то все равно имеем ненулевую погрешность (рис. 1). Такое происходит по описанным выше причинам из-за того, что на концах интервала вторая производная обращается в ноль и что функцию можно провести несколькими способами, но вмешивается коэффициент сглаживания. Восстановление функции по истинным интегралам, показанное на рисунке 1, происходит со средней относительной погрешностью 3.669414%. Сплайн строится таким образом, чтобы площадь под графиком приближалась к площади под каждой ступенькой, где ступенчатая линия показывает среднее значение функции на интервале  $I_i/(t_{i+1} - t_i)$ . Как можно заметить, наибольшая погрешность наблюдается на правом конце интервала. Этот результат показателен, даже при идеальных наблюдениях восстановление не идеально, что надо учесть при дальнейшем анализе.



**Рис. 1.** Восстановление функции по истинным значениям интегралов  $I_i$ . Ступенчатая линия  $avg$  – средние значения  $I_i/(t_{i+1} - t_i)$ , гладкая сплошная линия  $Spline$  – восстановленная функция, гладкая пунктирная линия  $f(t)$  – исходная функция

Теперь перейдем к результатам экспериментов. Цель данной статьи – исследование точности от неопределенности внутри процесса образования событий. Но в начале определим точность не от неопределенности внутри процесса, а от погрешности в самих наблюдениях.

### Эксперимент 1, погрешность в наблюдениях

Моделируя процесс потребления продукции так же, как было описано выше, получаем даты  $t_i$  и объемы  $y_i$  пополнения запасов, также определяем истинные значения интегралов  $I_i$  исходной функции между этими датами. Но восстанавливать значение функции будем по истинным интегралам, взятым с некоторой дополнительной относительной погрешностью  $y_i = I_i \cdot (1 + \varepsilon_i)$ . То есть наблюдениями будут значения, которые случайным образом будут чуть меньше или чуть больше истинных значений интегралов на случайное количество процентов. Величина  $\varepsilon_i$  будет генерироваться по нормальному закону распреде-

ления с нулевым математическим ожиданием и с заданным средним квадратичным отклонением в 1%, 5%, 10%, 20% и 30%. Так как случайные величины с нормальным законом распределения могут отклоняться на несколько сигм в большую или меньшую сторону, отдельные наблюдения могут достаточно серьезно отклоняться от истинных значений. Для каждого установленного значения СКО наблюдений проводим 20 независимых прогонов. Результаты эксперимента представлены в таблице 1.

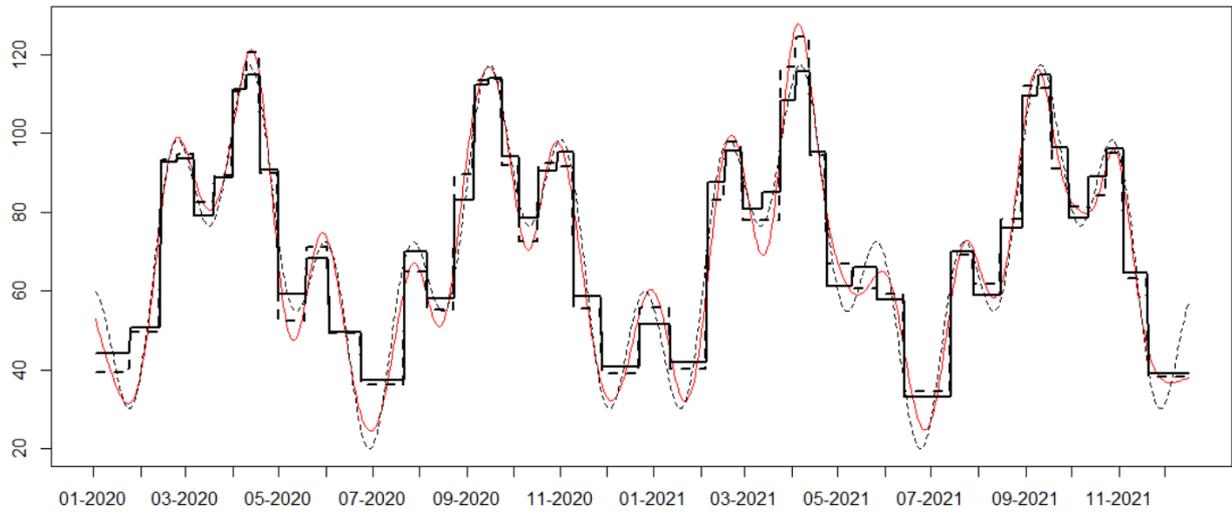
Таблица 1

СРЕДНЕЕ ОТНОСИТЕЛЬНОЕ ОТКЛОНЕНИЕ ВОССТАНОВЛЕННОЙ ФУНКЦИИ  
ОТ ФАКТИЧЕСКОГО СКО НАБЛЮДЕНИЙ

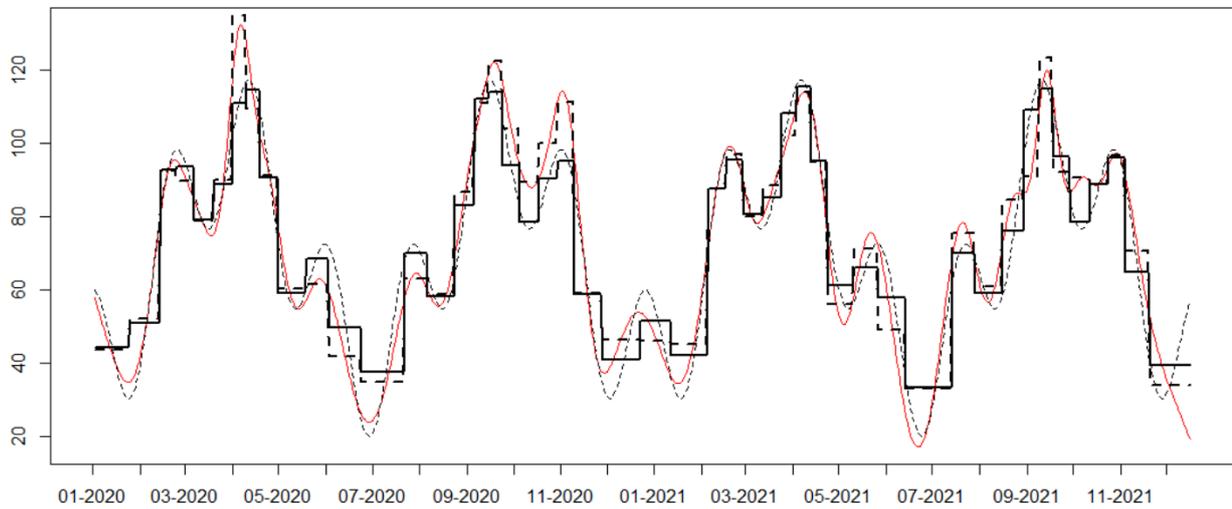
Прогоны	$\sigma_{\text{набл}} = 1\%$		$\sigma_{\text{набл}} = 5\%$		$\sigma_{\text{набл}} = 10\%$		$\sigma_{\text{набл}} = 20\%$		$\sigma_{\text{набл}} = 30\%$	
	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %
1	3.959	0.958	6.788	5.051	9.910	9.437	19.964	18.946	26.672	27.420
2	4.282	0.936	7.292	4.972	11.840	9.922	22.931	21.148	36.855	37.383
3	3.968	0.978	6.116	4.590	11.003	11.306	17.438	15.139	29.734	30.575
4	3.979	0.877	7.987	5.017	9.052	8.725	20.876	20.793	34.634	30.828
5	3.870	1.074	6.057	5.060	13.540	11.600	19.484	19.038	28.170	27.926
6	4.239	1.061	7.017	4.652	9.875	9.633	18.517	17.500	33.216	32.552
7	3.802	1.081	7.856	5.904	9.880	9.433	20.394	20.829	28.924	28.519
8	4.056	1.073	6.654	4.874	11.503	10.762	22.970	23.316	29.102	25.802
9	3.877	0.861	6.525	4.414	15.092	11.953	17.169	18.459	24.398	25.092
10	3.783	1.088	6.205	4.262	11.434	10.155	16.815	18.471	27.825	28.338
11	4.202	1.076	6.790	5.015	12.730	9.918	19.897	20.233	36.034	32.030
12	3.974	0.837	6.257	4.362	10.701	9.209	21.802	21.902	31.457	30.687
13	4.019	1.045	7.721	5.047	12.022	11.077	18.806	18.909	33.146	28.528
14	4.072	1.090	6.143	5.274	9.479	9.319	25.098	22.724	32.491	28.472
15	3.932	0.823	6.505	5.545	11.839	11.022	22.588	22.041	31.982	33.311
16	4.138	1.193	6.540	5.259	9.072	10.639	19.975	22.463	28.013	31.434
17	4.138	1.161	6.961	5.902	11.646	10.207	22.453	23.913	25.215	26.945
18	4.108	1.039	6.381	4.785	11.355	9.814	21.334	22.429	35.981	33.053
19	4.039	0.999	5.687	4.479	11.729	9.837	21.003	18.220	27.879	31.475
20	4.251	0.951	5.778	4.288	10.776	10.804	18.083	17.357	28.471	26.623

При погрешности наблюдений в 1% средняя относительная погрешность восстановления исходной функции колеблется от 3.8 до 4.2%, но не забываем, что даже при нулевой погрешности наблюдений погрешность восстановления функции была 3.669414%, поэтому увеличение погрешности наблюдений с 0 до 1% не такое значительное, как кажется. Также надо обратить внимание, что фактическое СКО наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$  не полностью объясняет погрешность восстановления исходной функции  $R$ . Когда несколько наблюдений подряд одновременно отклоняются в одну и ту же сторону, на этом участке восстановленная функция становится сильно смещена. На рисунке 2 показаны результаты некоторых прогонов восстановления функции по интегралам при заданных относительных погрешностях наблюдений  $\sigma_{\text{набл}}$ . На диаграмме рассеивания (рис. 3) заметно, что при сравнительно одних и тех же значениях  $\sigma_{\text{факт}}^{\text{набл}}$  точность восстановления исходной функции  $R$  может значительно отличаться. Визуально заметен практически линейный рост  $R$  от  $\sigma_{\text{факт}}^{\text{набл}}$ , но полиномиальная модель подходит чуть лучше.

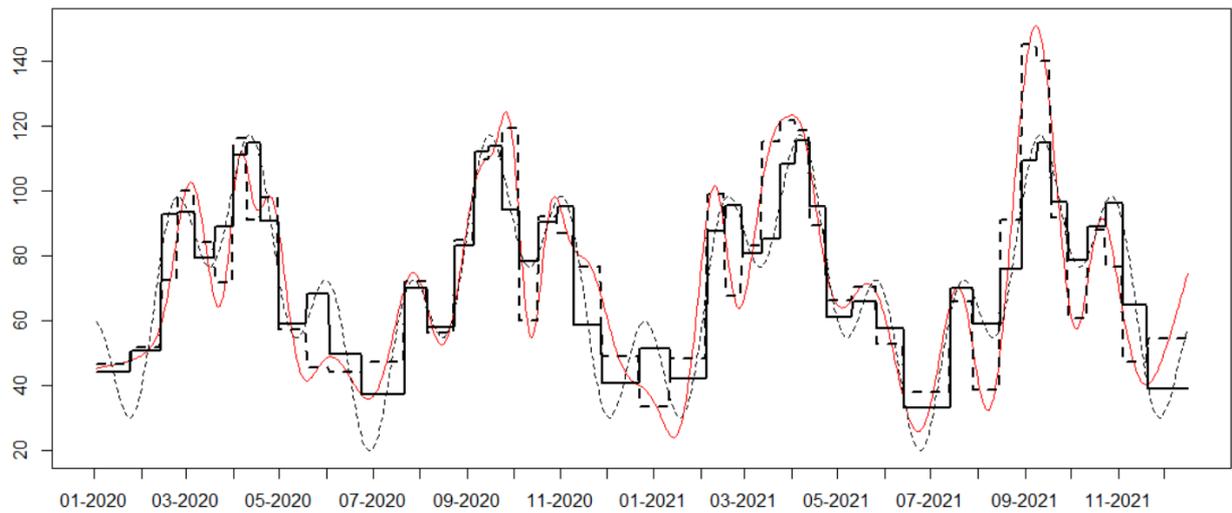
ИССЛЕДОВАНИЕ ТОЧНОСТИ ЕМКОСТНОГО МЕТОДА АНАЛИЗА РЕДКИХ СОБЫТИЙ  
 ОТ НЕОПРЕДЕЛЕННОСТИ ВНУТРИ ПРОЦЕССА ОБРАЗОВАНИЯ СОБЫТИЙ



(а) -  $\sigma_{\text{набл}} = 5\%$



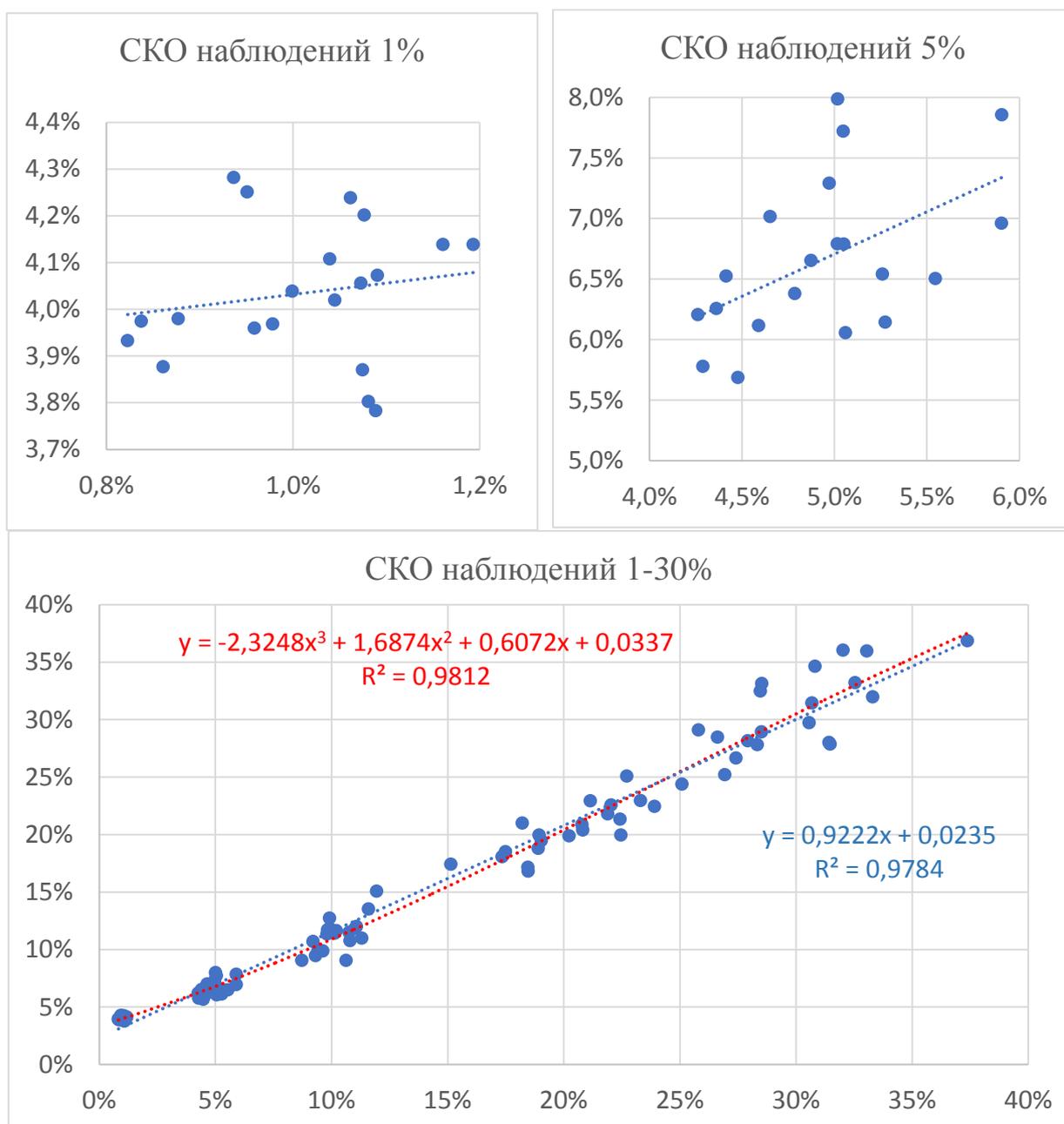
(б) -  $\sigma_{\text{набл}} = 10\%$



(в) -  $\sigma_{\text{набл}} = 20\%$

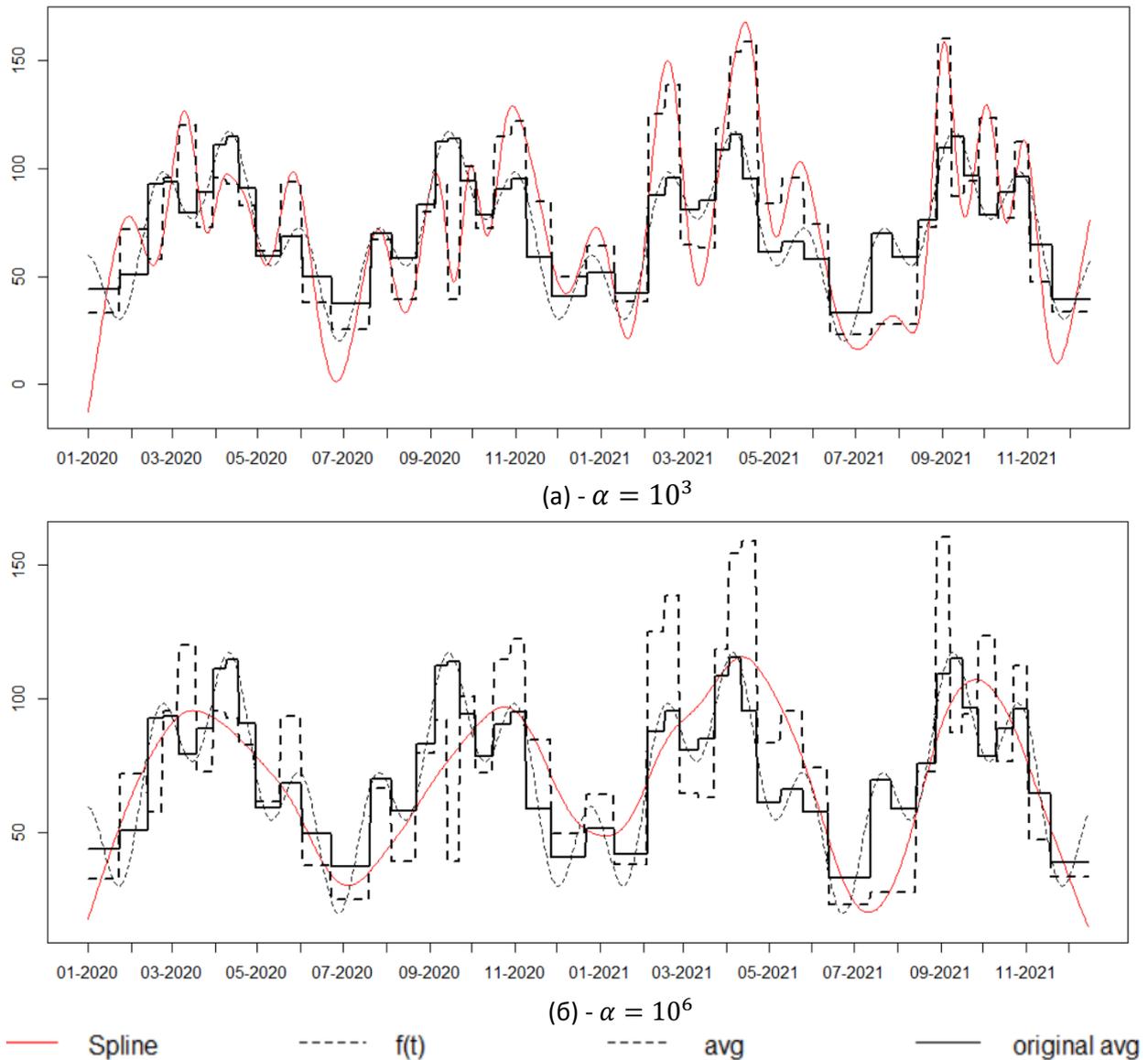
— Spline      - - - - f(t)      - - - - avg      — original avg

**Рис. 2.** Восстановление функции по наблюдениям, образованным от истинных значений интегралов с дополнительной относительной погрешностью  $y_i = I_i \cdot (1 + \varepsilon_i)$ : а – 5%, б – 10%; в – 20%



**Рис. 3.** Диаграмма рассеивания погрешности  $R$  восстановления исходной функции от фактического СКО наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$

При большом разбросе наблюдений от истинных значений, например, когда СКО относительной погрешности составляет 30%, появляются ложные дополнительные пики функции, точность восстановления функции неудовлетворительна (рис. 4а). В этом случае можно увеличить сглаживающий коэффициент  $\alpha$  и убрать высокочастотные пики (рис. 4б). Однако, конечно же, при этом теряются и высокочастотные составляющие исходной функции.



**Рис. 4.** Восстановление функции по интегралам при сильно зашумленных данных  $\sigma^{\text{набл}} = 30\%$ , удаление шумов увеличением коэффициента сглаживания: а)  $\alpha = 10^3$ ; б)  $\alpha = 10^6$

### Эксперимент 2, погрешность в процессе при проверке запасов

В данном эксперименте мы проверим, как влияет неопределенность внутри самого процесса образования событий на точность восстановления исходной функции. Напомним, что процессом образования событий является процесс потребления, когда из запасов ежедневно вычитается некоторый объем потребления, заданный определенной функцией, происходит проверка уровня запасов, если он опустился ниже критического уровня, то происходит мгновенное пополнение запаса. Предположим, что неопределенность кроется в процессе проверки уровня запасов, когда потребитель на глаз определяет, достаточно ли еще запасов или нет. Предположим, что во время такой проверки существует разброс, соответствующий нормальному распределению с нулевым математическим ожиданием и с определенным средним квадратичным отклонением, заданным относительно критического уровня в процентах. То есть пусть проверка необходимости пополнения запаса происходит по следующему условию (величина  $\varepsilon_i$  имеет нормальное распределение с нулевым математическим ожиданием и с заданным средним квадратичным отклонением  $\sigma^{\text{кр.ур.}}$ , которое будем изменять во время эксперимента):

$$Stock \leq CriticalStorage \cdot (1 + \varepsilon_i).$$

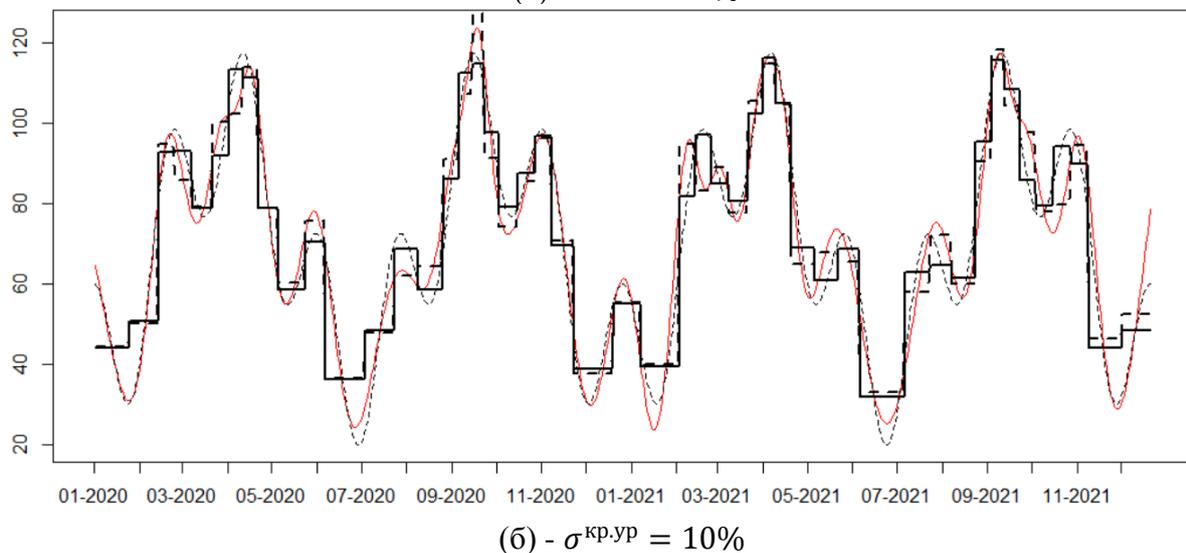
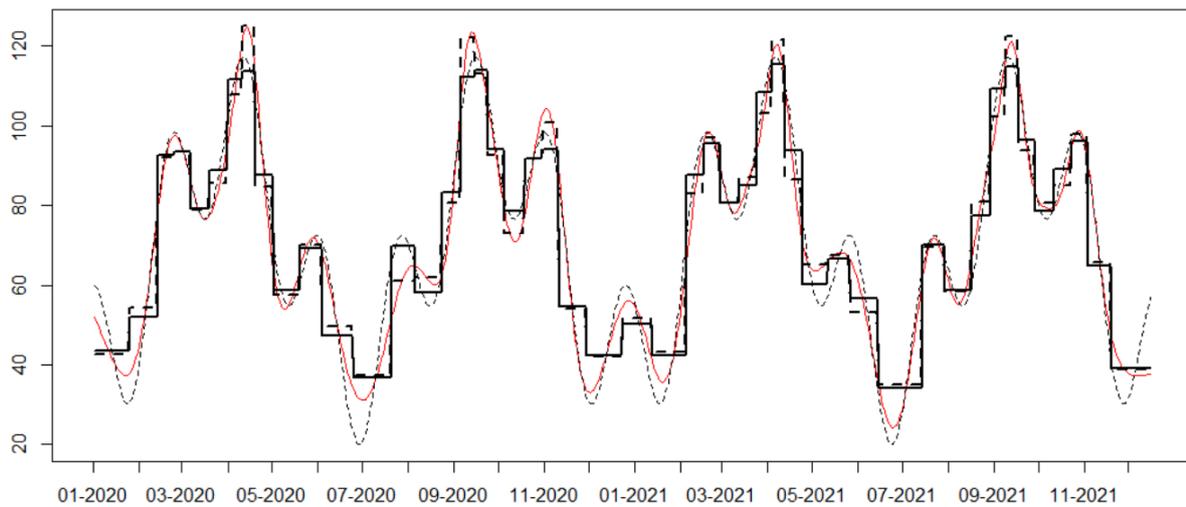
Стоит обратить внимание на то, что проверка уровня запасов во время моделирования происходит каждый день. Более вероятно, что условие будет выполняться чуть раньше, так как достаточно, чтобы лишь в один из дней до исходной даты пополнения запаса отклонение от критического уровня произошло на достаточную величину. Для того чтобы пополнение запаса произошло позже, необходимо, чтобы ни в один день до этого (спустя предыдущее пополнение) не было таких отклонений, вероятность чего меньше. Однако пусть объем пополнения запаса после такой проверки будет происходить как обычно, то есть без этого случайного отклонения, когда реальное количество запасов пополняется до максимума. В этом случае объемы пополнения запасов будут меньше, так как израсходовалось меньше запасов. Получается, что пополнение запасов будет происходить чуть раньше, но и объемы пополнения будут чуть меньше. Тем не менее по-прежнему будет возникать погрешность в наблюдениях  $\sigma_{\text{факт}}^{\text{набл}}$ , так как наблюдаемый объем пополнения запаса не будет совпадать с реально расходуемым количеством продукции. Причем величина этого несовпадения будет пропорциональна разбросу при сравнении критического уровня  $\sigma^{\text{кр.ур}}$  (данное обстоятельство тяжелее объяснить, связано это с тем, что наблюдаемый объем пополнения запаса хоть и связан с объемом расходуемой продукции, но он определяет, насколько хватит продукции до следующего события, а не до текущего, тем самым наблюдение и интеграл от функции с точки зрения расходуемой продукции относятся к разным интервалам времени, на каждом из которых происходит свой разброс). После того, как все нюансы проведения эксперимента обговорены, приведем его результаты (табл. 2).

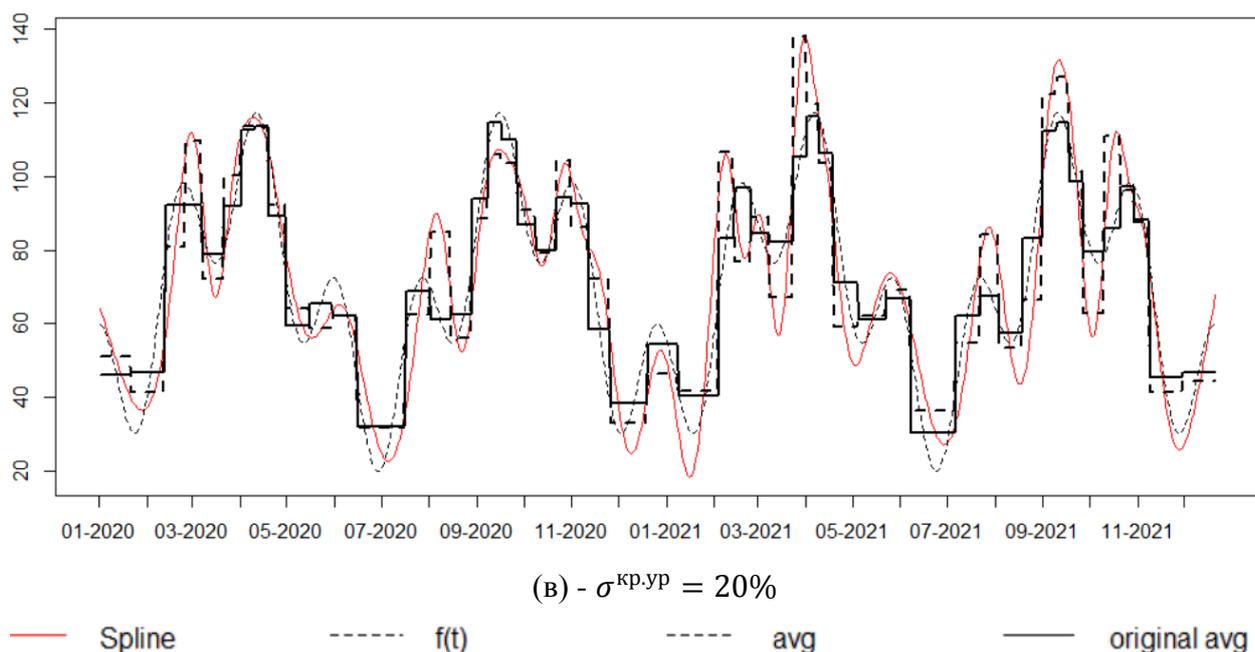
Таблица 2

СРЕДНЕЕ ОТНОСИТЕЛЬНОЕ ОТКЛОНЕНИЕ  $R$  ВОССТАНОВЛЕННОЙ ФУНКЦИИ И ФАКТИЧЕСКОЕ СКО НАБЛЮДЕНИЙ ОТ РАЗБРОСА ПРИ ПРОВЕРКЕ УРОВНЯ ЗАПАСОВ

Прогнозы	$\sigma^{\text{кр.ур}} = 1\%$		$\sigma^{\text{кр.ур}} = 5\%$		$\sigma^{\text{кр.ур}} = 10\%$		$\sigma^{\text{кр.ур}} = 20\%$		$\sigma^{\text{кр.ур}} = 30\%$	
	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %
1	5.119	3.328	7.013	4.656	7.147	6.724	14.697	14.174	18.513	17.680
2	4.760	3.150	6.816	4.629	6.799	6.819	13.656	12.829	12.919	13.582
3	4.597	2.891	5.710	4.823	7.067	6.010	15.285	13.025	14.493	18.531
4	5.651	3.333	5.458	3.909	7.082	7.200	12.395	10.708	10.700	14.327
5	4.951	3.177	5.697	4.477	7.769	7.526	12.782	12.548	15.001	15.634
6	4.344	3.070	4.016	3.674	8.069	6.693	11.344	11.790	22.514	23.348
7	4.822	2.996	5.964	4.884	8.738	6.683	10.317	9.388	16.715	16.533
8	5.145	3.041	7.191	4.819	10.765	8.350	10.708	9.872	14.745	15.777
9	5.021	3.070	7.413	5.384	7.519	6.931	11.564	12.579	17.315	18.146
10	4.162	3.282	6.216	3.844	7.388	5.768	12.022	10.240	15.773	15.853
11	5.021	3.070	5.723	4.758	9.086	8.520	11.418	10.640	14.162	17.326
12	4.042	2.890	5.336	4.885	8.574	7.041	10.760	11.293	21.328	20.465
13	5.001	3.282	5.392	4.182	7.157	6.474	11.010	11.285	13.156	18.410
14	5.005	3.142	5.659	4.677	7.433	6.857	12.183	12.236	18.570	18.755
15	4.659	3.544	6.027	5.647	7.344	6.209	14.498	15.300	14.670	15.075
16	4.329	2.950	7.044	5.382	9.920	7.101	11.117	13.105	18.975	21.918
17	5.431	3.623	5.163	4.355	9.011	6.987	12.381	11.508	21.095	21.712
18	5.200	3.378	6.145	4.381	10.019	8.974	10.794	10.135	23.465	23.128
19	5.316	3.702	7.537	4.939	8.426	7.030	15.109	14.028	19.108	20.920
20	5.475	3.345	5.808	4.123	7.980	8.346	12.120	12.623	26.281	29.147

Из результатов проведения экспериментов видно, что увеличение разброса погрешности при проверке уровня запасов приводит к тому, что увеличивается среднее квадратичное отклонение самих наблюдений, при этом оно растет медленнее. Объяснить это можно тем, что критический уровень запасов был меньше, чем средний объем пополнения запаса (который зависит от максимального уровня), поэтому отклонение от критического уровня вносит меньше погрешности, чем отклонения самих объемов пополнения запаса. Восстановление функции из наблюдений даже при разбросе в 20% при проверке уровня запасов происходит достаточно хорошо (рис. 5). Тем не менее можно констатировать, что неопределенность при проверке уровня запасов всего лишь влияет на погрешность самих наблюдений. Именно погрешность наблюдений влияет на погрешность восстановления исходной функции. Из диаграммы рассеивания (рис. б) видна очень похожая линейная зависимость, однако коэффициент наклона меньше (был 0.9222, стал 0.821, причем доверительные интервалы параметра наклона не пересекаются даже для доверительной вероятности 99%). Утверждать, что погрешность растет медленнее, чем в предыдущем эксперименте, не получится, так СКО наблюдений не говорит о том, как отклонения шли во время самого прогона. Так, два подряд отклонения в одну сторону сильнее заставят функцию отклониться, нежели если отклонения будут чередоваться то в одну сторону, то другую, а грубая оценка среднего квадратичного отклонения не способна это отобразить. Весьма вероятно, что при других прогонах, которые будут обладать той же оценкой СКО наблюдений, погрешность восстановления функции будет значительно отличаться.





**Рис. 5.** Восстановление функции по наблюдениям, образованным в процессе потребления с погрешностью при сравнении уровня запасов  $Stock \leq CriticalStorage \cdot (1 + \varepsilon_i)$ : : а – 5%, б – 10%; в – 20%



**Рис. 6.** Диаграмма рассеивания погрешности  $R$  восстановления исходной функции от фактического СКО наблюдений  $\sigma_{факт}^{набл}$  при разбросе во время проверки уровня запасов

### Эксперимент 3, погрешность в дате события

В предыдущем эксперименте разброс при проверке критического уровня приводил к сдвигу даты события, как правило, в одну сторону (событие возникало раньше). В этом эксперименте мы проверим, как влияет разброс в дате возникновения события на точность

восстановления исходной функции. Под разбросом даты возникновения события можно представить такую картину, когда потребитель идет пополнять запасы чуть позже или чуть раньше положенного срока. Одной из причин такого может быть по-прежнему то, что потребитель не может точно оценить свои запасы. Другой причиной может быть то, что потребитель занят в определенный день и планирует пополнить запасы либо заранее, либо чуть позже, когда освободится.

Для проведения эксперимента внесем следующие изменения в модель. Вначале определяется дата  $t_i$  завершения запаса так же, как и в первом эксперименте, при условии, что уровень запасов опустился ниже критического уровня (без разброса). Затем от данной даты откладываем разброс, выраженный через относительное значение интервала времени от предыдущего события  $\Delta t_i = t_i - t_{i-1}$ :

$$\hat{t}_i = t_i + \Delta t_i \cdot \varepsilon_i.$$

Случайная величина  $\varepsilon_i$  по-прежнему будет соответствовать нормальному закону с нулевым математическим ожиданием и с заданным средним квадратичным отклонением  $\sigma^{\Delta t_i}$ . Однако из-за того, что интервалы времени между событиями  $\Delta t_i$  не одинаковые, разброс  $\Delta t_i \cdot \varepsilon_i$  на самом деле соответствует нестационарному Гауссовскому процессу. Также, так как моделирование происходит по дням, пусть разброс происходит в целых днях, а не в дробных, то есть будем округлять разброс до ближайшего целого. Так как интервалы между событиями не такие большие (колеблются от 9 до 28 дней), отклонение в 1% с последующим округлением ничего не дает. Начинать проводить эксперименты сразу будем с  $\sigma^{\Delta t_i} = 5\%$ . При этом мы будем определять фактическое относительное среднее квадратичное отклонение даты события  $\sigma_{\text{факт}}^{\Delta t_i}$  в процентах, которое считаем уже после всех округлений.

В определенный таким образом момент времени наступления события  $\hat{t}_i$  наш потребитель будет пополнять запасы, причем пополнять будет те запасы, которые он израсходует к этому моменту  $\hat{t}_i$  (то есть к сдвинутому), а не к первоначальному (не сдвинутому)  $t_i$ . Таким образом, если потребитель чуть раньше решит пополнять запасы, то и объем пополнения запасов будет меньше, а если он решит позже пополнять запасы, то объем пополнения запасов будет больше. Расчет следующего события будет происходить именно с этого нового момента времени  $\hat{t}_i$  (сдвинутого), то есть при моделировании придется откатить модельное время немного назад, если разброс времени события был отрицательный, и продвинуть немного вперед, если разброс был положительным. Хотя объемы пополнения запаса будут меняться из-за сдвига, все равно будет наблюдаться разброс в самих наблюдениях по сравнению с истинными интегралами от исходной функции (истинные интегралы определяются также до сдвинутой даты  $\hat{t}_i$ ), так как этот объем пополнения определяет дату следующего события, а не текущего. В каждом прогоне будем, как и раньше, определять фактическое среднее квадратичное отклонение самих наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$  в процентах. Результаты моделирования представлены в таблице 3, а результаты некоторых прогонов изображены на рисунке 7.

Из результатов эксперимента видно, что разброс относительно даты события вызывает достаточно значительный разброс в самих наблюдениях. Так, при  $\sigma_{\text{факт}}^{\Delta t_i} = 10.958\%$  разброс в наблюдениях составил 16.868%, что в свою очередь привело к погрешности восстановления функции  $R = 16.640\%$ . Объяснить это можно тем, что несовпадение

наблюдаемых объемов пополнения запаса и реально расходуемых объемов обладает удвоенной дисперсией, что должно приводить к тому, что разброс наблюдений должен быть в среднем на 1.41 больше (из-за того, что исходная функция нестационарная, разброс может быть еще больше).

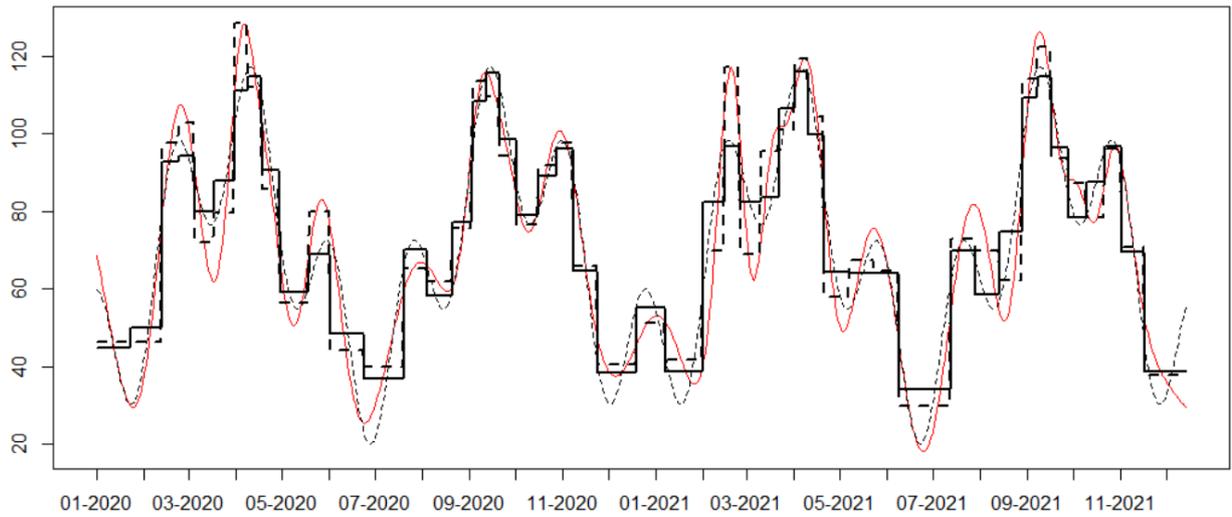
Таблица 3

СРЕДНЕЕ ОТНОСИТЕЛЬНОЕ ОТКЛОНЕНИЕ  $R$  ВОССТАНОВЛЕННОЙ ФУНКЦИИ,  
ФАКТИЧЕСКОЕ СКО НАБЛЮДЕНИЙ И ФАКТИЧЕСКОЕ СКО ДАТЫ СОБЫТИЙ  
ОТ РАЗБРОСА В ДАТЕ ОБРАЗОВАНИЯ СОБЫТИЯ

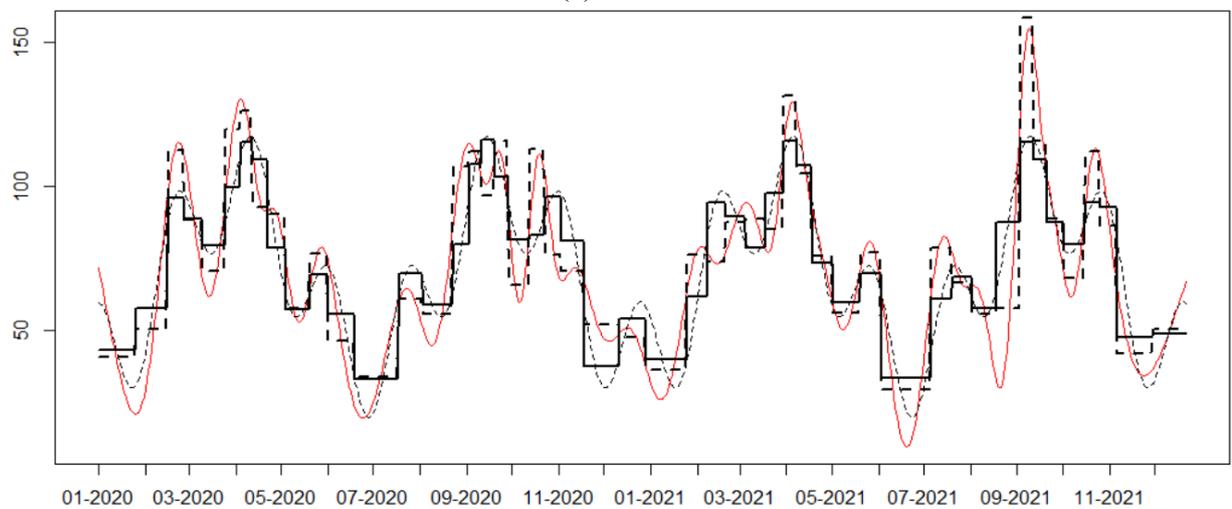
Прогоны	$\sigma^{\Delta t_i} = 5\%$			$\sigma^{\Delta t_i} = 10\%$			$\sigma^{\Delta t_i} = 20\%$		
	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$\sigma_{\text{факт}}^{\Delta t_i}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$\sigma_{\text{факт}}^{\Delta t_i}$ в %	$R$ в %	$\sigma_{\text{факт}}^{\text{набл}}$ в %	$\sigma_{\text{факт}}^{\Delta t_i}$ в %
1	9.853	9.141	5.052	16.640	16.868	10.958	37.770	35.168	21.582
2	9.126	7.807	5.201	13.224	11.810	9.148	27.999	30.582	26.842
3	10.005	8.517	5.367	24.391	21.434	11.343	27.815	30.156	23.606
4	9.965	8.765	5.501	11.024	11.373	8.927	20.885	26.269	25.831
5	7.582	6.600	5.499	13.917	11.356	8.708	35.196	32.752	21.257
6	8.870	8.169	6.025	14.191	13.167	8.342	31.770	30.089	20.668
7	10.763	8.781	5.561	14.027	14.455	11.014	32.355	34.865	23.430
8	10.255	8.751	5.510	16.057	16.328	10.957	29.163	35.353	28.833
9	8.535	7.689	5.299	16.992	18.160	10.351	36.468	35.337	29.040
10	10.041	8.956	5.987	16.472	16.408	10.357	20.734	24.238	18.118
11	11.086	9.107	5.465	19.672	16.569	9.745	31.335	27.144	17.385
12	11.124	7.541	3.994	16.727	14.809	10.175	33.706	32.106	22.201
13	12.396	9.692	5.386	13.665	12.372	8.469	22.821	23.309	16.438
14	8.396	7.555	5.111	16.178	15.666	10.980	30.813	37.947	28.022
15	9.722	6.794	5.373	14.807	13.972	8.648	35.335	28.307	19.815
16	9.092	8.904	6.516	15.220	16.387	11.979	25.431	28.997	22.877
17	10.021	9.311	5.587	19.917	17.846	13.230	24.071	23.455	17.580
18	10.372	9.748	6.104	19.292	20.349	11.942	31.596	27.755	19.910
19	12.656	13.668	7.212	13.052	11.982	8.731	33.720	37.302	22.554
20	10.056	9.371	5.274	16.158	15.017	9.408	29.485	26.171	17.962

Еще хочется обратить внимание на то, что сам по себе разброс даты событий иногда приводит к тому, что событие возникает абсурдно рано. Так, например, при разбросе в  $\sigma^{\Delta t_i} = 20\%$ , если отклонение произойдет на 3 сигмы, то произойдет сдвиг на 60% от интервала времени, то есть от предыдущего события прошло лишь 40% от положенного времени. Получается, что очередное событие возникло в 2.5 раза раньше, такое значительное отклонение будет тяжело назвать случайным, его легко спутать с каким-то структурным фактором. Дальнейшее увеличение разброса даты событий, например  $\sigma^{\Delta t_i} = 30\%$  или 40%, вообще может привести к тому, что очередное событие появится через нулевой период времени или должно появиться до появления текущего. Большие отклонения в датах события маловероятны, на практике такое трудно представить.

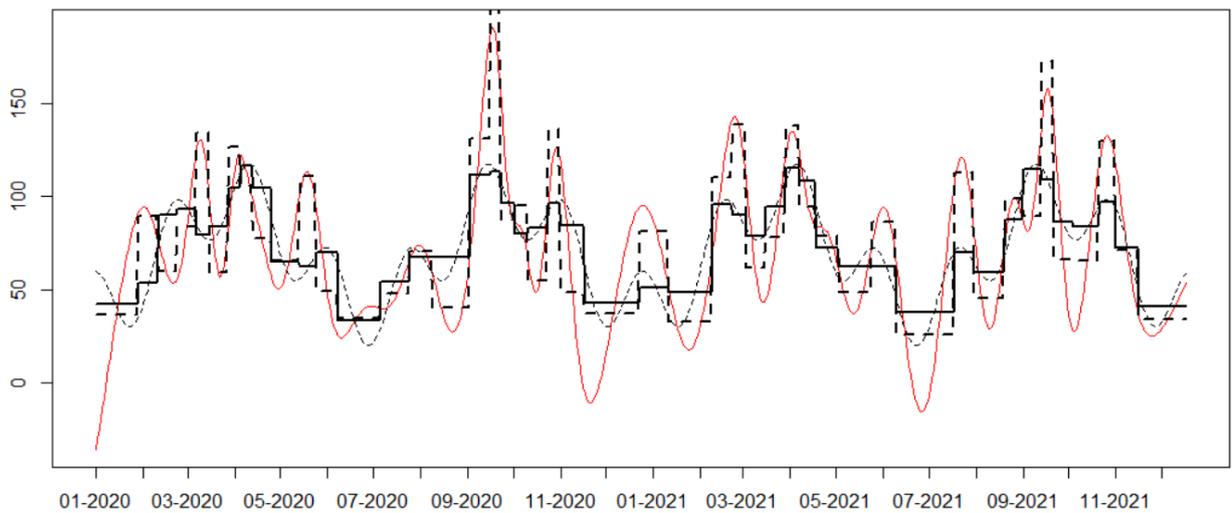
ИССЛЕДОВАНИЕ ТОЧНОСТИ ЕМКОСТНОГО МЕТОДА АНАЛИЗА РЕДКИХ СОБЫТИЙ  
ОТ НЕОПРЕДЕЛЕННОСТИ ВНУТРИ ПРОЦЕССА ОБРАЗОВАНИЯ СОБЫТИЙ



(a) -  $\sigma^{\Delta t_i} = 5\%$



(б) -  $\sigma^{\Delta t_i} = 10\%$

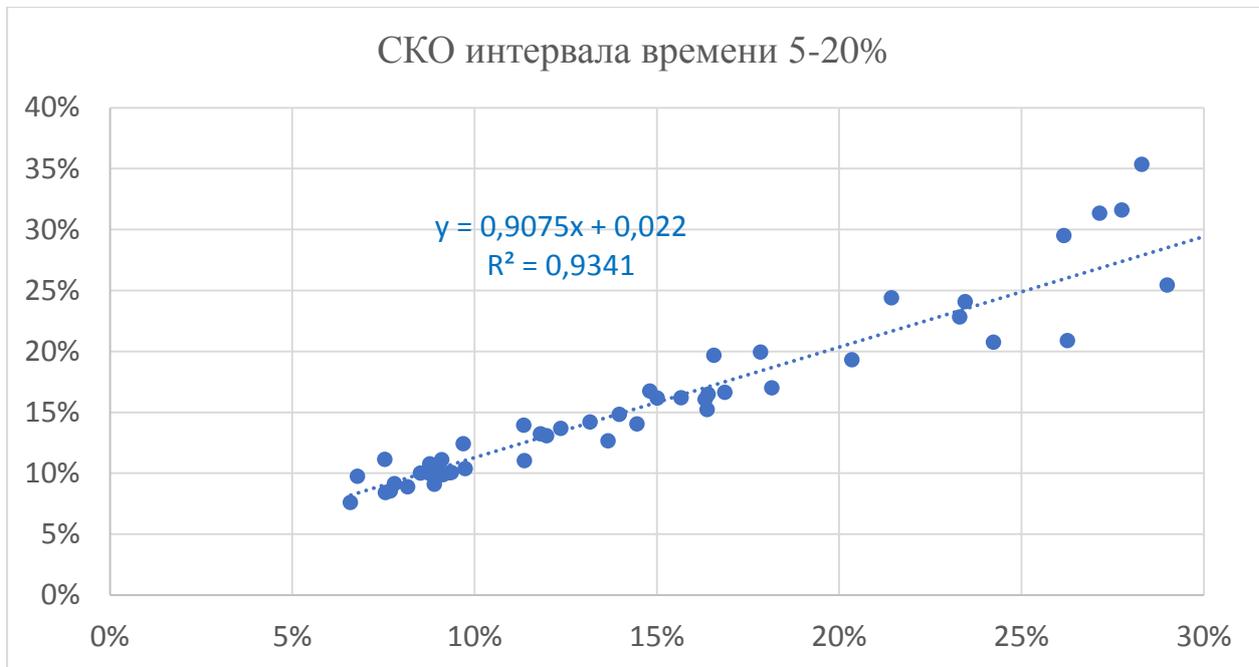


(в) -  $\sigma^{\Delta t_i} = 20\%$

— Spline      - - - - - f(t)      - - - - - avg      — original avg

**Рис. 7.** Восстановление функции по наблюдениям, образованным в процессе потребления при разбросе в дате образования события с СКО интервала времени  $\sigma^{\Delta t_i}$ : а – 5%, б – 10%; в – 20%

На диаграмме рассеивания, изображенной на рисунке 8, видна линейная зависимость погрешности восстановления функции от фактической погрешности наблюдений, коэффициент наклона получился равен 0.907 (в первом эксперименте был 0.9222, во втором – 0.821). Доверительный интервал для коэффициента наклона в этом эксперименте пересекается с доверительными интервалами в предыдущих двух экспериментах.

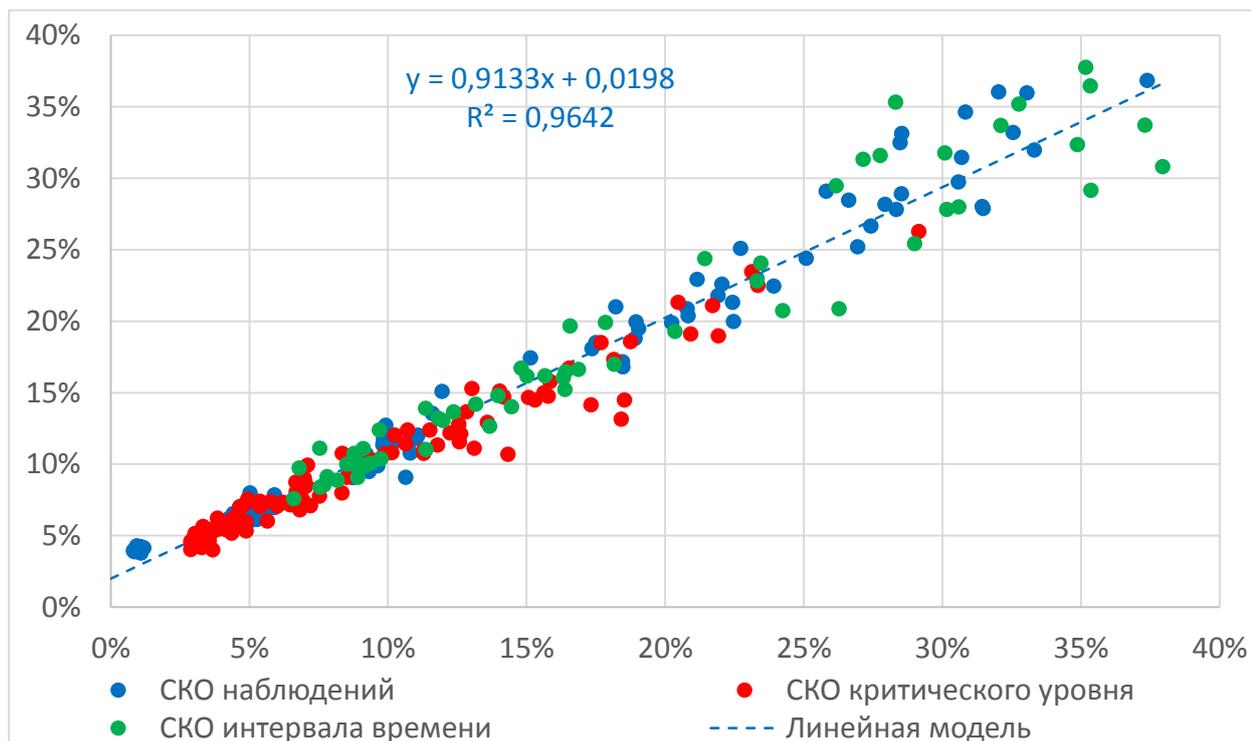


**Рис. 8.** Диаграмма рассеивания погрешности  $R$  восстановления исходной функции от фактического СКО наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$  при разбросе даты события

На рисунке 9 представлена диаграмма рассеивания для данных всех трех экспериментов, объединенных вместе. Результаты каждого эксперимента обозначены своим цветом. Наблюдается линейная зависимость относительной погрешности восстановления исходной функции  $R$  от фактического среднеквадратичного отклонения наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$ . Получившаяся линейная зависимость имеет уравнение

$$R = 0.913316 \cdot \sigma_{\text{факт}}^{\text{набл}} + 0.019825$$

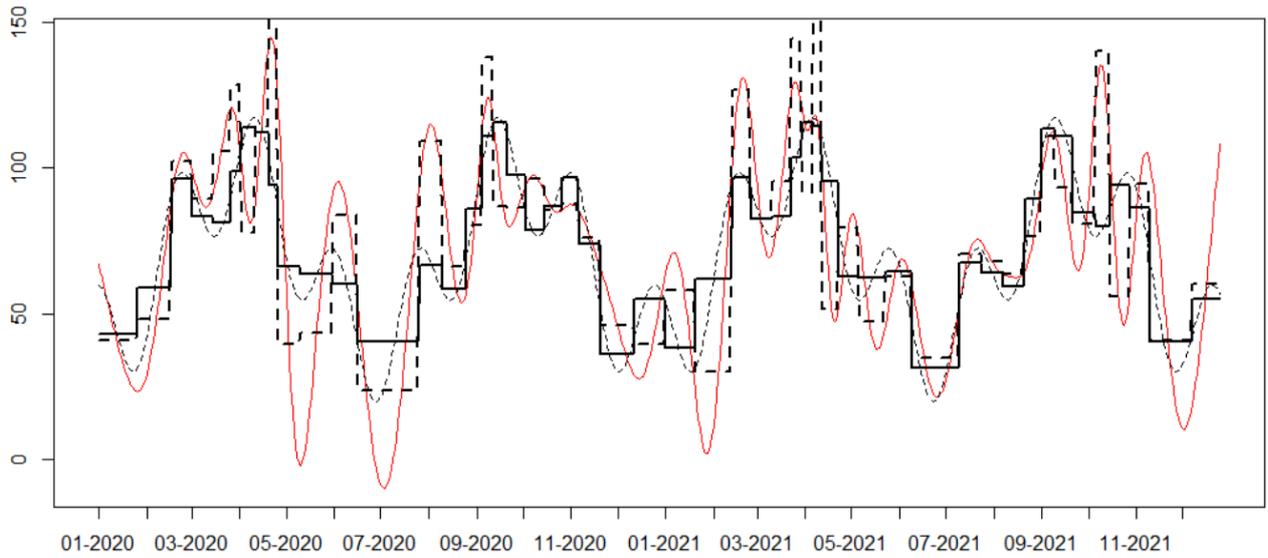
(0.010959)                      (0.001747)



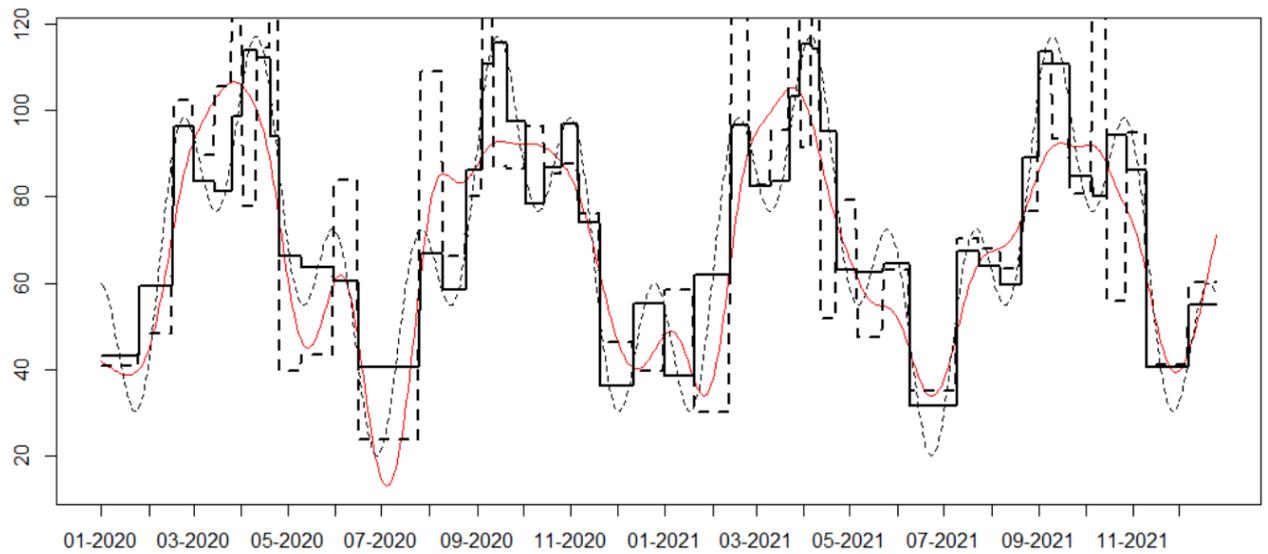
**Рис. 9.** Диаграмма рассеивания погрешности  $R$  восстановления исходной функции от фактического СКО наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$  для всех трех экспериментов

#### ЗАКЛЮЧЕНИЕ

По результатам трех экспериментов этой части исследования мы видим, что погрешность внутри самого процесса образования событий приводит к появлению погрешности в наблюдениях и не более того. Точность восстановления исходной функции по последовательности интегралов зависит линейно от этой погрешности в наблюдаемых интегралах. При погрешности наблюдений  $\sigma_{\text{факт}}^{\text{набл}}$  около 5% восстановление функции происходит очень хорошо, восстановленная функция повторяет все изгибы исходной функции, экстремумы приблизительно находятся на соответствующем месте, амплитуда отклонений не очень большая. При погрешности  $\sigma_{\text{факт}}^{\text{набл}}$  около 10% по-прежнему восстановление происходит достаточно хорошо, однако амплитуда отклонения восстановленной функции становится заметно больше, также экстремумы могут быть смещены более сильно. При погрешности  $\sigma_{\text{факт}}^{\text{набл}}$  в 20 и 30% начинают появляться ложные экстремумы там, где их не должно быть. То есть на восстановленную функцию как будто накладывается дополнительный шум из случайных отклонений. Данный шум можно попробовать сгладить с помощью использования более высокого значения коэффициента сглаживания в процессе восстановления функции по интегралам. На рисунке 10, в рамках последнего эксперимента, изображен пример удаления этих нежелательных шумов при помощи увеличения коэффициента сглаживания с  $\alpha = 10^3$  до  $\alpha = 10^5$ . Однако при удалении высокочастотных шумов также теряются высокочастотные колебания самой исходной функции. Этот подход можно использовать при сильно зашумленных данных, когда интерес представляет восстановление по интегралам хотя бы основных гармоник исходной функции.



(a) -  $\alpha = 10^3$



(б) -  $\alpha = 10^5$

— Spline      - - - - - f(t)      - - - - - avg      — original avg

**Рис. 10.** Восстановление функции по интегралам при разбросе в дате образования события с СКО интервала времени  $\sigma^{\Delta t_i} = 20\%$  (фактическая СКО погрешности наблюдений  $\sigma_{\text{факт}}^{\text{набл}} = 30.518\%$ ) и удаление шумов увеличением коэффициента сглаживания:  
 а)  $\alpha = 10^3$ ; б)  $\alpha = 10^5$

За рамками этой части исследования остались другие интересные нас моменты. Хотелось бы посмотреть, как разброс других параметров процесса потребления влияет на точность восстановления функции. Такими параметрами могут быть максимальный уровень запасов, периодичность проверки запасов, неопределенность самой исходной функции и т.д.

## ЛИТЕРАТУРА

1. *Кораблев Ю.А.* Емкостный метод определения функции скорости потребления // Экономика и менеджмент систем управления. Воронеж: Изд-во «Научная книга», 2015. Т. 15. № 1.1. С. 140-150.
2. *Кораблев Ю.А., Голованова П.С., Кострица Т.А.* Емкостный метод анализа редких событий в торговле различными товарами // Бизнес. Образование. Право. Вестник Волгоградского института бизнеса. 2019. № 3(48). С. 121-131. DOI: 10.25683/VOLBI.2019.48.313.
3. *Кораблев Ю.А.* Метод восстановления функции по интегралам для анализа и прогнозирования редких событий в экономике // Экономика и математические методы. М.: ЦЭМИ РАН, 2020. Т. 56. № 3. С. 113-124.
4. *Кораблев Ю.А.* Погрешность емкостного метода анализа редких событий, удаленность от конечного потребителя // Известия Кабардино-Балкарского научного центра РАН. 2019. № 3 (89). С. 48-77. DOI: 10.35330/1991-6639-2019-3-89-48-77.
5. *Green P.J., Silverman B.W.* Nonparametric regression and generalized linear models. A roughness penalty approach. New York: Chapman & Hall, 1994.
6. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. Изд. 2-е. М.: Наука. Главная редакция физико-математической литературы, 1979.

## REFERENCE

1. Korablev Yu.A. *Yemkostnyy metod opredeleniya funktsii skorosti potrebleniya* [Capacity method of the determination of consumption rate function] // *Ekonomika i menedzhment sistem upravleniya* [Economics and management of control systems]. Voronezh: Izd-vo «Nauchnaya kniga – Scientific Book Print House». 2015. V. 15. № 1.1. Pp. 140-150.
2. Korablev Yu.A., Golovanova P.S., Kostritsa T.A. *Yemkostnyy metod analiza redkikh sobytiy v torgovle razlichnymi tovarami* [Capacity method of analyzing rare events in the trade of various goods] // *Biznes. Obrazovaniye. Pravo* [Business. Education. Law]. *Vestnik Volgogradskogo instituta biznesa* [Bulletin of the Volgograd Institute of Business]. 2019. № 3. Pp. 121–131. DOI: 10.25683/VOLBI.2019.48.313
3. Korablev Yu.A. *Metod vosstanovleniya funktsii po integralam dlya analiza i prognozirovaniya redkikh sobytiy v ekonomike* [The function restoration method by integrals for analysis and forecasting of rare events in the economy] // *Ekonomika i matematicheskiye metody* [Economics and Mathematical Methods]. Moscow: TSEMI RAN, 2020. Vol. 56. № 3. Pp. 113-124.
4. Korablev Yu.A. *Pogreshnost' yemkostnogo metoda analiza redkikh sobytiy, udalennost' ot konechnogo potrebitelya* [Error of the capacity method of rare events analysis, remoteness from the end user] // *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [“The News of KBSC of RAS”]. 2019. № 3 (89). С. 48-77. DOI: 10.35330/1991-6639-2019-3-89-48-77
5. Green P.J., Silverman B.W. Nonparametric regression and generalized linear models. A roughness penalty approach. New York: Chapman & Hall, 1994.
6. Tikhonov A.N., Arsenin V.Ya. *Metody resheniya nekorrektnykh zadach* [Methods for solving ill-posed problems]. Ed. 2nd. M.: Science. The main editorial office of physical and mathematical literature, 1979.

# ACCURACY RESEARCH OF THE CAPACITY METHOD FOR ANALYZING RARE EVENTS FROM UNCERTAINTY WITHIN THE PROCESS OF EVENT FORMATION\* (Part 1)

**Yu.A. KORABLEV**

Financial University under the Government of the Russian Federation (Finuniversity)  
125993, Russia, Moscow, Leningradsky prospect, 49  
Email: academy@fa.ru

*The study determines the influence of uncertainty within the consumption process of the rare events formation on the accuracy of restoring the original consumption function from the rare events data using the capacity method. The function is restored from a sequence of integrals observed with an error using a cubic integral smoothing basis spline. In the first experiment in this part of the study, the influence of the error in the observations on the accuracy of the function reconstruction is determined. In the second experiment, the influence of the random spread during the check of the stock level is determined, and in the third experiment, the influence of the spread in the dates of the events on the actual observation error and on the error in restoring the original function is determined.*

*Experiments have shown that the uncertainty within the process of the events formation affects only on the error in observations, which affects linearly on the error of the restored function. A model of this linear relationship is built. At the same time, a large observation error leads to the appearance of noise on the restored function. A mechanism for removing this noise by increasing the smoothing coefficient in the process of restoring the function from integrals is proposed.*

**Keywords:** rare events, capacitive method, consumption process, consumption rate, recovery, regression, integral spline, error, accuracy.

*Работа поступила 01.09.2020 г.*

## **Сведения об авторе:**

**Кораблев Юрий Александрович**, к.э.н., доцент кафедры «Системный анализ в экономике» Финансового университета при Правительстве Российской Федерации (Финуниверситет).

125993, г. Москва, Ленинградский проспект, 49.

Тел.: 8-916-882-72-43.

E-mail: yura-korablyov@yandex.ru

## **Information about the author:**

**Korablev Yuri Alexandrovich**, Candidate of Economics, Associate Professor, Financial University under the Government of the Russian Federation (Finuniversity), Department of System Analysis in Economics.

125993, Moscow, Leningradsky prospect, 49.

Ph.: 8-916-882-72-43.

E-mail: yura-korablyov@yandex.ru

---

\* The study was carried out with the financial support of the Russian Foundation for Basic Research within the framework of scientific project No. 19-010-00154