УДК 330.42(045), 51-77(045)

MSC: 62H30; 62P20; 91B42

DOI: 10.35330/1991-6639-2020-3-95-74-91

МОДЕЛИРОВАНИЕ И ВОССТАНОВЛЕНИЕ ПРЕДПОЧТЕНИЙ ПОКУПАТЕЛЕЙ МЕЖДУ ДВУМЯ АЛЬТЕРНАТИВНЫМИ ТОВАРАМИ С ПОМОЩЬЮ ЕМКОСТНОГО МЕТОДА АНАЛИЗА РЕДКИХ СОБЫТИЙ В ЭКОНОМИКЕ* (Часть 1)

Ю.А. КОРАБЛЕВ

Финансовый университет при Правительстве Российской Федерации (Финуниверситет)

125993, г. Москва, Ленинградский проспект, 49

E-mail: academy@fa.ru

В статье предложен подход для восстановления динамики изменения предпочтения потребителя между двумя альтернативными товарами из данных редких событий, связанных с покупками этих товаров, основанный на емкостном методе анализа редких событий. Кратко изложена основная идея емкостного метода, в котором редкие события анализируются с точки зрения процессов, происходящих внутри источников событий. Процессы потребления, которые являются наиболее распространенными в экономике при образовании событий, можно моделировать как процессы опустошения/наполнения емкости. Это позволяет восстанавливать скорость изменения уровня этой емкости (скорость изменения запаса продукции) с помощью математического метода восстановления функции по интегралам. Рассмотрены четыре варианта образования событий при пополнении запасов двух альтернативных товаров. Для каждого варианта даны предложения, как можно восстановить динамику изменения предпочтения потребителя между двумя альтернативными товарами. Для варианта, в котором потребитель пополняет запасы обоих товаров, когда заканчивается запас любого из товара, необходимо предварительно выполнить задачу распознавания и определить, образовано ли событие завершением первого или второго товара. Даны рекомендации для осуществления этого распознавания. Показаны примеры восстановления динамики изменения предпочтения между двумя альтернативными товарами.

Ключевые слова: редкие события, емкостный метод, скорость потребления, восстановление, предпочтение, альтернативные товары.

Анализ данных имеет большое и важное значение в современных исследованиях в любой области, поэтому развитие новых методов анализа данных является актуальной задачей. Особняком стоят методы исследования редких событий, для которых не применимы классические статистические методы на основе временных рядов. Для анализа данных редких событий применяют специальные методы.

Часть таких методов по-прежнему рассматривает данные редких событий через призму временных рядов, в результате чего получаются ряды, содержащие большое количество нулевых значений. Но в этих методах производят особенные манипуляции, чтобы «корректно» обрабатывать эти ряды с нулевыми и ненулевыми значениями. Например, метод Кростона [1, 2] предполагает разбить временной ряд на два ряда – ряд ненулевых значений и ряд с длительностью между ненулевыми значениями. Далее по методу Кростона предполагается экспоненциальное сглаживание полученных двух рядов и предположение, что следующее событие возникнет через ожидаемое (сглаженное) количество нулевых интервалов со значением, равным ожидаемому (сглаженному) ненулевому значению. Одна-

* Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета 2020 года

ко данный метод самый примитивный, ненулевое значение не обязательно появляется через заданное количество нулевых значений, само ненулевое значение может изменяться, экспоненциальное сглаживание не способно находить закономерности.

Другим подходом обработки данных редких событий в виде временных рядов, содержащих большое количество нулевых значений, являются методы классификации. Например, метод ближайших соседей [3, 4] просматривает ряд данных, двигая по ряду данных скользящее окно фиксированной ширины, и пытается определить участки (соседей), которые наиболее похожи на последовательность последних наблюдений, ориентируясь на некоторую норму, например, евклидово расстояние. Затем для всех наиболее похожих соседей рассматривается следующее за ними наблюдение, подсчитывается либо среднее арифметическое значение, либо вычисляется взвешенное значение на основе той самой «похожести» (евклидово расстояние), которое затем возвращается как прогнозное будущее значение. Причем в данном случае для распознавания используются признаки, которые построены как вектор прошлых значений (скользящее окно, которое движется по временному ряду).

Очень часто в качестве признаков используются не данные самого временного ряда, полученного из данных редких событий, а внешние признаки, которые наблюдаются одновременно со значениями временного ряда. Самым популярным методом классификации в данном случае является метод логистической регрессии [5], который работает с данными, представленными в виде нулей и единиц (относится ли это событие к одному классу или к другому). Задача метода тогда заключается в классификации по имеющемуся входному набору признаков складывающейся обстановки и прогнозировании (на основе классификации) редкого события в период, к которому отнесены эти наблюдаемые признаки. То есть, например, если перед каким-то редким явлением наблюдаемые признаки имеют тенденцию показывать определенные значения, то когда мы вновь наблюдаем похожие значения признаков, это свидетельствует о большом шансе возникновения этого явления. Методы классификации очень популярны для анализа редких событий, но у них есть свои недостатки. Так, из-за того, что в обучающей выборке доля положительных наблюдений крайне мала, оценка вероятности возникновения этих событий оказывается смещенной (заниженной) и приходится корректировать ее особым образом. Существенным же недостатком является то, что некоторым редким событиям не предшествуют никакие наблюдаемые характерные изменения признаков, признаки изменяются не до возникновения события, а во время или после наступления его. Также хочется обратить внимание, что задача классификации или распознавания немного отличается от задачи анализа данных. В задачах распознавания мы ищем знакомые образы, в то время как в задаче анализа данных мы пытаемся выявить закономерности.

Еще одним подходом работы с редкими событиями, представленными временным рядом, является метод Виллемейна (бутстреппинга) [6, 7], идея которого заключается в случайном отборе фиксированного количества наблюдений, подсчете суммы и повторении этой процедуры множество раз, в результате чего можно искусственно построить эмпирическую функцию распределения для суммы заданного количества случайно отобранных наблюдений. Этот метод используют в логистике для определения запасов на таком уровне, чтобы этих запасов хватило на заданное количество будущих периодов времени для удовлетворения спроса с заданной доверительной вероятностью. Данный метод также слишком наивный, исходное распределение значений во временном ряду может быть неслучайным или значения будут коррелированными, а слепой отбор этих значений с помощью равномерного распределения не сможет этого учесть.

Совсем по-другому на редкие события смотрят в теории случайных процессов [8], в которой данные о редких событиях представляются в виде Пуассоновских потоков или потоков Пальма, в которых предполагается, что время между событиями является случайной величиной с некоторым законом распределения. В этом случае по имеющимся данным определяют характеристики закона распределения. Для Пуассоновского потока время между событиями соответствует экспоненциальному закону с нестационарной интенсивностью событий, которую определяют как функцию от времени. Для потока Пальма закон распределения времени между событиями может быть произвольным, но он всегда стационарен, для потока Пальма определяют вид распределения и его стационарные параметры. С помощью Пуассоновских потоков можно найти вероятность появления заданного количества событий за некоторое время, но при этом нельзя определить моменты времени возникновения событий. С помощью потоков Пальма можно определить моменты времени следующих событий, отложив математическое ожидание интервалов между событиями от предыдущего события, однако так как закон распределения стационарен, будущие события будут идти с одинаковым шагом, что не всегда соответствует действительности.

Конечно же, в некоторых условиях для определенных целей перечисленные методы могут быть достаточны, однако для целей прогнозирования как моментов возникновения событий, так и значений этих событий описанные методы либо не применимы, либо обладают большой погрешностью из-за неспособности выявлять закономерности. Поэтому разработка новых методов анализа редких событий является актуальной задачей.

ИДЕЯ ЕМКОСТНОГО МЕТОДА

В разработанном мной методе анализа редких событий представление событий ближе всего к потокам событий, когда события имеют дискретный характер. Но при этом я утверждаю, что их появление происходит не через случайные периоды времени, а подчинено определенным процессам, которые происходят в источниках образования событий, и целью анализа является определение этих процессов и их параметров. Для анализа событий первым делом надо отделить одни события от других, ни в коем случае не перемешивать и не агрегировать данные событий, образованные в разных источниках. Для этого всего лишь надо отказаться от агрегирования событий и построения из редких событий временных рядов, а разделять события по идентификаторам источников. Следующим шагом является предположение на основе всей имеющейся информации вида процесса, в результате которого образуются события. Процесс может быть задан как некоторый алгоритм, в результате работы которого формируются события. Затем из имеющихся данных редких событий необходимо определить параметры данного процесса, то есть осуществить восстановление (регрессию) параметров процесса. Для этого надо получить обратный алгоритм, который бы из дискретных событий мог определить значения параметров процесса (если в математике обратную функцию к $y(x)$ изображают иногда как $y^{-1} = x(y)$, то для алгоритма функционирования процесса мы получаем как бы алгоритм в минус первой степени, который из результатов работы процесса получает параметры процесса). После того, как параметры процесса были восстановлены, нужно произвести поиск закономерности изменения этих параметров с ходом времени или от изменения внешних условий, если такие изменения параметров наблюдаются. Целью этого этапа является экстраполяция параметров процесса на будущее для того, чтобы с этими параметрами на последнем этапе запустить процесс образования событий и получить прогноз будущих дискретных событий.

В экономике самым распространенным процессом образования событий является процесс потребления, который работает по следующей схеме. С ходом времени постоянно уменьшается запас некоторой продукции, и по достижению некоторого критического уровня срабатывает событие, которое пополняет запас продукции на определенную величину. Другим таким процессом может являться процесс накопления возмущения до некоторого уровня, после которого срабатывает событие и возмущение сбрасывается. Оба этих процесса можно моделировать как процессы опустошения или наполнения емкости, метод исследования таких процессов я называю емкостным [8-12]. Параметром процесса будет являться нестационарная скорость опустошения/наполнения емкости $f(t)$, которую можно восстановить из данных редких событий. Такой функцией может являться спрос от времени, индивидуальная скорость потребления продукции, интенсивность покупок у выбранного покупателя, интенсивность накопления возмущения (у источника событий, а не у наблюдателя). Для восстановления этой функции инвертируем процесс потребления продукции, получаем обратную задачу к задаче управления запасами, когда по имеющимся данным (t_i, y_i) – моментах времени и величинах воздействия события (покупок) – определяется скорость $f(t)$ изменения запаса неподконтрольных нам покупателей. Для этого 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} , т.е. $y_i = \int_{t_i}^{t_{i+1}} f(t)dt$. Это предположение справедливо для процессов потребления, оно негласно используется в логистике в системах управления запасами [13], когда строится пилообразный график во время моделирования собственных запасов. Однако на практике по многим причинам наблюдаемые значения y_i в точности не совпадают со значениями интегралов, всегда присутствует некоторая погрешность ε_i . Тогда предположение будет выглядеть следующим образом: $y_i = \int_{t_i}^{t_{i+1}} f(t)dt + \varepsilon_i$.

Задача восстановления функции по ее интегралам превращается в оптимизационную задачу, в которой минимизируется сумма квадратов между наблюдаемыми значениями и интегралами от функции плюс дополнительный штраф на нелинейность (C – произвольная большая константа, влияющая на сглаживание; n – номер последнего события).

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

Решение данной задачи получено мной в виде интегрального кубического сглаживающего базисного сплайна со штрафной функцией [14], оно не входит в планы данной статьи, но будет многократно здесь использоваться.

Использование емкостного метода позволяет чуть ли не с идеальной точностью восстанавливать исходную функцию, заложенную при моделировании тех же систем управления запасами, в которых события образуются в результате процессов, схожих с процессами опустошения или наполнения емкости. В результате получается восстановить скорость (интенсивность) расходования запаса продукции из наблюдений редких событий, провести поиск закономерностей, экстраполировать их и получить прогноз будущих покупок неподконтрольного нам покупателя (получить прогноз будущих событий, возникающих в процессе потребления). Полученный результат считаю законченным. Но хочется пойти дальше и использовать данный метод в новых, совсем других задачах. Одно из применений, которое мне удалось обнаружить, связано с восстановлением изменения предпочтения клиента между двумя альтернативными товарами по данным покупок данного товара.

ЗАДАЧА ВОССТАНОВЛЕНИЯ ПРЕДПОЧТЕНИЯ МЕЖДУ ДВУМЯ АЛЬТЕРНАТИВНЫМИ ТОВАРАМИ

Существует множество научных работ, посвященных моделированию предпочтений, ранжированию товаров в соответствии с этими предпочтениями и принятию решений в зависимости от предпочтений. Сами же предпочтения иногда связывают с функцией полезности. Разработано множество математических способов, как моделировать функцию полезности для лица, принимающего решение [15]. Однако среди всего множества научных публикаций практически нет работ, посвященных выявлению этих самых предпочтений из данных приобретения различных товаров. Все имеющиеся публикации используют либо маркетинговые методы, такие как соцопросы и анкетирование, которые взаимодействуют с живыми людьми и выявляют их мнение о разных товарах (например, работы [16, 17]), либо методы исследования социальных сетей, что по сути те же методы опроса населения, но которые не уведомляют людей о том, что их мнение собирается, так как те сами добровольно все о себе рассказывают. В одной из последних работ [18] используется метод анализа настроений по текстам сообщений и нечеткие множества, чтобы обрабатывать неточности при этом анализе. Однако методов, которые выявляют предпочтения среди товаров, а тем более динамику изменения этих предпочтений со временем, найти не удалось. Предлагаемый подход будет таким средством выявления предпочтений и может быть интересен для маркетологов.

Говоря об альтернативных товарах, я подразумеваю такие товары, которые удовлетворяют одну и ту же потребность потребителя, но при этом по некоторым причинам потребитель не хочет потреблять лишь один из товаров, а хочет разнообразить их. Например, это могут быть продукты питания, белый и черный хлеб или гречневая и геркулесовая крупа. То есть такие товары, которые могут заменить друг друга. Для простоты изложения и большего понимания иногда буду давать физическую интерпретацию, в которой участвуют такие товары.

Прежде всего попробуем представить задачу восстановления предпочтения между двумя альтернативными товарами на формализованном языке математики и логики. Пусть есть потребитель, который пополняет запасы каждого товара. Товары считаются равнозначными для удовлетворения некоторой общей для обоих товаров потребности. Пусть каждый товар удовлетворяет общую потребность в одинаковом объеме. Причем сама эта общая потребность может также меняться со временем. Обозначим потребность как функцию от времени $f(t)$. Такой потребностью может быть, например, потребность в заданном количестве килокалорий каждый день, причем сама эта потребность может быть неизвестна (не наблюдаться). Потребитель для удовлетворения этой потребности использует как первый, так и второй товар, причем в зависимости от предпочтений он может использовать один товар больше, чем другой товар. Предпочтения также могут меняться со временем. Предпочтение первого товара обозначим как функцию от времени $\alpha(t)$, а предпочтение второго товара – $\beta(t) = 1 - \alpha(t)$, где функция $\alpha(t) \in [0; 1]$. То есть предпочтение будет показывать долю использования товара в общем использовании обоих товаров (функции $\alpha(t)$ и $\beta(t)$ безразмерные). Функции $\alpha(t)$ и $\beta(t)$ так же, как и $f(t)$, напрямую не наблюдаются, функции $\alpha(t)$ и $\beta(t)$ могут быть независимы от функции потребности $f(t)$. В результате скорость потребления первого товара составит $f^1(t) = \alpha(t)f(t)$, а скорость потребления второго товара – $f^2(t) = \beta(t)f(t) = (1 - \alpha(t))f(t)$.

У рассматриваемого потребителя есть некоторые запасы первого и второго товара, которые расходуются со скоростью $f^1(t) = \alpha(t)f(t)$ и $f^2(t) = \beta(t)f(t) = (1 - \alpha(t))f(t)$ соответственно. Когда запасы товаров подходят к концу, данный потребитель вынужден по-

полнить запасы. В результате образуется набор данных объемов покупок y_i^1, y_i^2 первого и второго в моменты времени t_i . Однако тут есть несколько вариантов, как происходит расход продукции и пополнение запаса в зависимости от схемы поведения потребителя. Рассмотрим несколько вариантов.

1). Когда любой из товаров закончился, потребитель совершает покупку только того товара, который закончился, и не покупает тот товар, который еще не закончился (например, в момент t_i покупается первый товар в объеме y_i^1 , когда второй товар не покупается $y_i^2 = 0$). В данной схеме поведения получается так, что события, которые образовались в результате пополнения запаса первого товара, совершенно не зависят от событий, которые образовались в результате пополнения второго товара. Всю имеющуюся выборку наблюдений можно разбить на две независимые выборки (t_i, y_i^1) и (t_i, y_i^2) (индекс i в разных выборках разный, независимый). Так, объемы покупок каждого товара не зависят от объемов покупок другого товара, мы имеем ситуацию, когда наблюдаемые объемы покупок каждого товара расходуются от момента совершения покупки до момента времени совершения следующей покупки. То есть в рамках этого предположения получается так, что объем каждой покупки есть интеграл от скорости расхода за время между текущим и следующим событием, имеем $y_i^1 = \int_{t_i}^{t_{i+1}} f^1(t) dt$ и $y_i^2 = \int_{t_i}^{t_{i+1}} f^2(t) dt$. По этим данным можно с легкостью восстановить функции скорости расхода продукции $\hat{f}^1(t) = \hat{\alpha}(t)\hat{f}(t)$ и $\hat{f}^2(t) = \hat{\beta}(t)\hat{f}(t)$, воспользовавшись упомянутым выше методом восстановления функции по интегралам. Тогда предпочтения между товарами определяются следующим образом: $\hat{\alpha}(t) = \hat{f}^1(t)/(\hat{f}^1(t) + \hat{f}^2(t))$ и $\hat{\beta}(t) = \hat{f}^2(t)/(\hat{f}^1(t) + \hat{f}^2(t))$.

2). Когда запасы одного товара заканчиваются, потребитель начинает потреблять запасы другого товара до тех пор, пока и другой товар не закончится, а когда оба товара заканчиваются, потребитель совершает покупку и пополняет оба товара одновременно. То есть тут возникает только одно событие, привязанное к моменту времени, когда запасы обоих товаров закончились. Получается так, что за время между двумя событиями расходуется суммарный объем обоих покупок, и этот объем соответствует интегралу от общей функции потребности: $y_i^1 + y_i^2 = \int_{t_i}^{t_{i+1}} (f^1(t) + f^2(t)) dt = \int_{t_i}^{t_{i+1}} f(t) dt$. Из данных наблюдений $(t_i, y_i^1 + y_i^2)$ можно будет восстановить только общую функцию потребности $\hat{f}(t)$, но нельзя выявить предпочтения между товарами, так как никакой информации о скорости расхода каждого товара наружу не поступает. Может сложиться впечатление, что информация о предпочтении заключается в покупках y_i^1 и y_i^2 , раз мы наблюдаем эти объемы покупок, но это не так. Дело в том, что в данной схеме объем покупки одного товара может использоваться как замещение предпочтения в другом товаре. Например, если первый товар закончился раньше, то оставшаяся часть второго товара расходуется для компенсации расхода первого товара (часть от y_i^2 пошла на $\int_{t_i+\tau}^{t_{i+1}} f^1(t) dt$, где τ – время, за которое закончился первый товар). То есть потребитель был вынужден потреблять второй товар вместо первого, хотя ему этого не хотелось. В этой схеме не получится восстановить истинные предпочтения потребителя $\alpha(t)$ и $\beta(t)$.

3). Потребитель пополняет запасы не тогда, когда они заканчиваются, а в произвольный другой момент, когда ему по некоторым причинам оказалось удобно это сделать. Например, потребитель не специально вышел из дома, чтобы пополнить запасы, а возвращаясь с работы, на обратном пути по дороге домой зашел в магазин. То есть событие пополнения

запаса совсем не связано с моментами израсходования запаса. Без дополнительных предположений выяснить, как расходовался запас каждого товара, не получится. Однако если предположить, что покупатель всегда закупается до определенного уровня, можно сказать, что объем совершенной покупки есть объем потребления спустя предыдущее событие до текущего. То есть если покупатель всегда дома держит 10 буханок белого хлеба и 5 черного и сегодня он у вас купил 4 буханки белого и 1 черного, то значит, с момента времени предыдущей покупки он потребил ровно столько же буханок белого и черного хлеба. В этом случае получается, что наблюдаемый объем покупок и есть интеграл от скорости потребления с предыдущего события до текущего события, $y_i^1 = \int_{t_{i-1}}^{t_i} f^1(t) dt$ и $y_i^2 = \int_{t_{i-1}}^{t_i} f^2(t) dt$ (когда событие образуется вследствие завершения запаса, было $y_i^1 = \int_{t_i}^{t_{i+1}} f^1(t) dt$ и $y_i^2 = \int_{t_i}^{t_{i+1}} f^2(t) dt$, обратите внимание на индексы). При этом, если предположить, что такие события пополнения запаса всегда происходят раньше, чем запас любого из товаров заканчивается (каждый день человек, возвращаясь с работы, пополняет запас, не давая ему закончиться), то потребление одного товара оказывается независимым от потребления другого товара (человек потребляет только то, что ему в данный момент больше хочется, не заставляя себя есть то, что ему не очень хочется). Тогда мы будем иметь по-прежнему два независимых набора данных (t_i, y_{i+1}^1) и (t_i, y_{i+1}^2) (обратите внимание на индекс). Из этого набора данных мы по-прежнему можем восстановить скорость расхода продукции $\hat{f}^1(t) = \hat{\alpha}(t)\hat{f}(t)$ и $\hat{f}^2(t) = \hat{\beta}(t)\hat{f}(t)$ с помощью метода восстановления функции по интегралам. Еще раз подчеркну, что за время между событиями $t_{i+1} - t_i$ расходуется объем y_{i+1} , а не y_i . поэтому надо в исходных данных сдвинуть наблюдаемые объемы покупок на одну позицию. В результате у нас так же, как и в первом случае, получается восстановить функции предпочтений $\hat{\alpha}(t)$ и $\hat{\beta}(t)$. Конечно же, дополнительные предположения о том, что покупатель всегда пополняет запасы до фиксированного постоянного уровня, и предположение о том, что он делает это достаточно часто, чтобы запасы не закончились, могут быть неверны. Нарушение предположений требует дополнительных исследований.

4). Когда запасы одного товара заканчиваются, потребитель пополняет запасы как первого, так и второго товара. Событие пополнения запаса обоих товаров происходит при условии, что либо один товар закончился, либо другой товар закончился. При такой схеме поведения в момент срабатывания события один товар полностью израсходовался, когда другой товар может быть еще в наличии, но при этом его запас также пополняется. Также можно сказать, что срабатывание события для одного товара влечет появление события для второго товара. При такой схеме мы располагаем одним набором данных (t_i, y_i^1, y_i^2) , когда в один и тот же момент времени t_i наблюдаются покупки обоих товаров в объемах y_i^1 и y_i^2 соответственно. Данная схема поведения клиента более естественная и поэтому более интересная. С одной стороны, может показаться, что получится использовать тот же прием, что и в третьем варианте, когда события никак не были связаны с пополнением запаса, и можно с помощью сдвига данных на одну позицию произвести восстановление обеих функций предпочтения. Действительно, для того товара, который остался в наличии, пусть для первого, время события t_i не связано с расходом товара, и объем покупки y_i^1 (при предположении, что потребитель придерживается постоянного уровня) будет соответствовать объему потребленной продукции спустя предыдущее событие ($y_i^1 = \int_{t_{i-1}}^{t_i} f^1(t) dt$). Однако так как

для другого товара, который закончился, пусть для второго, событие привязано к моменту завершения запаса и за время спустя предыдущее событие израсходовался объем не текущей покупки y_i^2 , а предшествующей y_{i-1}^2 (т.е. $y_i^2 = \int_{t_i}^{t_{i+1}} f^2(t) dt$), сдвиг данных делать не нужно, также не нужно делать предположение, что покупатель пополняет запасы всегда до одного и того же уровня. Оказывается, что если правильно распознать среди набора данных, для какого из товаров произошло событие пополнения запаса, то получится по-прежнему корректно восстановить функцию изменения предпочтения потребителя. Рассмотрим этот вариант поведения потребителя более подробно.

МОДЕЛЬ ВОССТАНОВЛЕНИЯ ПРЕДПОЧТЕНИЯ МЕЖДУ ДВУМЯ АЛЬТЕРНАТИВНЫМИ ТОВАРАМИ

Чтобы подробно разобраться в этом случае, построим модель, в которой в искусственно заданных условиях действует потребитель, расходующий запасы первого и второго товара с разной скоростью в зависимости от изменяющихся предпочтений, затем проанализируем полученные данные на возможность восстановления этого предпочтения из данных покупок этого потребителя. Пусть будет задана исходная функция потребности $f(t)$, которая изменяется со временем определенным образом, пусть это будет гармоническая функция. Зададим исходные функции предпочтения $\alpha(t)$ и $\beta(t) = 1 - \alpha(t)$, пусть они также будут гармоническими функциями, но частота которых выше, чем у $f(t)$. Внешний вид функций $f(t)$ и $\alpha(t)$ изображен на рисунке 1.

Рис. 1. Функция потребности $f(t)$ (шт./в день) и функция предпочтения $\alpha(t)$ (в долях)

Далее строим модель на основе модели управления запасами, но при этом с двумя товарами, причем так, что когда один товар заканчивается, то пополняются запасы обоих товаров. Пусть пока что для простоты реализации запасы товаров пополняются до некоторого постоянного уровня, но при этом будем иметь в виду, что в общем виде это может не выполняться. Запасы пополняются тогда, когда запасы одного из товаров опустились ниже некоторого критического уровня (ниже нуля, после чего используются страховые запасы). Реализация модели в Excel представлена на рисунке 2. Подробно останавливаться на том, как устроена модель, не будем. В результате работы модели получаем моменты времени и объемы пополнения запасов продукции обоих товаров, данные представлены в таблице 1 (для большей точности не округляем значения до целых чисел). Рассмотрим эти данные подробно.

МОДЕЛИРОВАНИЕ И ВОССТАНОВЛЕНИЕ ПРЕДПОЧТЕНИЙ ПОКУПАТЕЛЕЙ МЕЖДУ ДВУМЯ АЛЬТЕРНАТИВНЫМИ ТОВАРАМИ С ПОМОЩЬЮ ЕМКОСТНОГО МЕТОДА АНАЛИЗА РЕДКИХ СОБЫТИЙ В ЭКОНОМИКЕ

	D	E	F	G	H	I	J	K	L	M	N	O
1	Оба товара пополняются одновременно						Товар 1	Товар 1				
2			Предпочтение			Макс запас	700	500				
3			одного товара			Начальный запас %	0,9	0,4				
4	Время	Потребление	над другим			Запас	=J2*J3			Запас	=K2*K3	
5	t	f(t)	alfa(t)	1-alfa(t)		630				200		
6	01.01.2020	130	0,7	0,3		539				161		
7	02.01.2020	129,9955552	0,699733363	0,300267		448,037773				121,9667		
8	03.01.2020	=ЕСЛИ((I5-E6*F6>0)*(M5-E6*G6>0);I5-E6*F6;\$J\$2)				357,1887573				82,83347		
9	04.01.2020	=ЕСЛИ((I5-E6*F6>0)*(M5-E6*G6>0);"-";D6)				266,528069				43,53415		
10	05.01.	=ЕСЛИ((I5-E6*F6>0)*(M5-E6*G6>0);"-"; \$J\$2-I5+E6*F6)				176,1302883				4,003021		
11	06.01.2020	129,8889452	0,693369563	0,30663		700	06.01.2020	613,9308		500	06.01.2020	535,8249
12	07.01.2020	129,8401246	0,690475515	0,309524		610,3485731				459,8113		
13	08.01.2020	129,7824617	0,68707359	0,312926		521,1784712				419,1989		
14	09.01.2020	129,7159736	=ЕСЛИ((I5-E6*F6>0)*(M5-E6*G6>0);M5-E6*G6;\$K\$2)							378,1014		
15	10.01.2020	129,64068	0,678783	=ЕСЛИ((I5-E6*F6>0)*(M5-E6*G6>0);"-";D6)						336,4587		
16	11.01.2020	129,556603	=ЕСЛИ((I5-E6*F6>0)*(M5-E6*G6>0);"-";\$K\$2-M5+E6*G6)							294,2126		
17	12.01.2020	129,4637682	0,668588307	0,331412		170,6935832				251,3068		
18	13.01.2020	129,3622024	0,662809219	0,337191		84,95112293				207,6871		
19	14.01.2020	129,251936	0,656596021	0,343404		0,084816104				163,3014		
20	15.01.2020	129,1330016	0,64996528	0,350035		700	15.01.2020	783,8472		500	15.01.2020	381,8996
21	16.01.2020	129,0054344	0,642934677	0,357065		617,0579327				453,9366		
22	17.01.2020	128,8692723	0,635522958	0,364477		535,1585516				406,9667		
23	18.01.2020	128,7245557	0,627749884	0,37225		454,3517266				359,049		
24	19.01.2020	128,5713273	0,619636183	0,380364		374,6842802				310,1451		
25	20.01.2020	128,4096326	0,611203487	0,388797		296,1998649				260,2199		
26	21.01.2020	128,2395195	0,602474282	0,397526		218,9388524				209,2414		
27	22.01.2020	128,0610385	0,593471843	0,406528		142,9382318				157,181		
28	23.01.2020	127,8742423	0,584220174	0,41578		68,2315197				104,0135		
29	24.01.2020	127,6791865	0,574743943	0,425256		700	24.01.2020	705,1513		500	24.01.2020	450,2829
30	25.01.2020	127,4759286	0,565068417	0,434932		627,9673788				444,5567		
31	26.01.2020	127,2645291	0,555219395	0,444781		557,307644				387,9519		
32	27.01.2020	127,0450505	0,545223137	0,454777		488,039743				330,1747		
33	28.01.2020	126,8175579	0,535106298	0,464894		420,178869				271,2181		
34	29.01.2020	126,5821186	0,524895853	0,475104		353,7364399				211,0784		
35	30.01.2020	126,3388025	0,514619026	0,485381		288,7200884				149,7559		
36	31.01.2020	126,0876817	0,504303219	0,495697		225,1336646				87,25467		
37	01.02.2020	125,8288305	0,493975939	0,506024		162,9772499				23,58225		
38	02.02.2020	125,5623256	0,483664721	0,516335		700	02.02.2020	597,7528		500	02.02.2020	541,25
39	03.02.2020	125,2882461	0,473397059	0,526603		640,6889128				434,0228		
40	04.02.2020	125,0066732	0,46320033	0,5368		582,7857806				366,9193		

Рис. 2. Пример модели в Excel

Таблица 1

ДАННЫЕ МОДЕЛИРОВАНИЯ, МОМЕНТЫ ВРЕМЕНИ И ОБЪЕМЫ ПОПОЛНЕНИЯ
ЗАПАСОВ ПРОДУКЦИИ ОБОИХ ТОВАРОВ

i	Дата t_i	Объем покупки		i	Дата t_i	Объем покупки		i	Дата t_i	Объем покупки	
		y_i^1	y_i^2			y_i^1	y_i^2			y_i^1	y_i^2
1	06.01.2020	613,93	535,82	39	11.01.2021	709,77	328,05	77	15.01.2022	777,14	387,77
2	15.01.2020	783,85	381,90	40	20.01.2021	735,60	423,96	78	25.01.2022	764,90	516,27
3	24.01.2020	705,15	450,28	41	29.01.2021	635,37	509,77	79	03.02.2022	572,27	562,28
4	02.02.2020	597,75	541,25	42	06.02.2021	469,11	531,92	80	11.02.2022	415,01	574,03
5	10.02.2020	436,13	557,89	43	13.02.2021	341,06	518,68	81	18.02.2022	302,30	545,14
6	17.02.2020	316,88	535,62	44	20.02.2021	289,10	553,12	82	25.02.2022	262,76	565,66
7	24.02.2020	272,10	561,96	45	27.02.2021	255,07	567,56	83	04.03.2022	243,27	564,26
8	02.03.2020	246,72	566,96	46	06.03.2021	241,51	559,75	84	11.03.2022	244,17	540,92
9	09.03.2020	241,94	549,70	47	13.03.2021	247,87	530,56	85	19.03.2022	302,72	565,47
10	16.03.2020	255,94	512,34	48	21.03.2021	312,05	548,19	86	28.03.2022	391,94	546,23
11	24.03.2020	327,85	520,32	49	30.03.2021	405,77	523,19	87	08.04.2022	563,87	526,20

12	03.04.2020	478,26	535,92	50	11.04.2021	636,43	537,30	88	21.04.2022	758,33	450,41
13	16.04.2020	726,36	515,24	51	23.04.2021	711,13	389,65	89	03.05.2022	721,98	321,10
14	28.04.2020	723,11	348,34	52	05.05.2021	717,82	314,22	90	16.05.2022	718,88	340,34
15	11.05.2020	753,43	332,97	53	18.05.2021	702,04	346,80	91	01.06.2022	701,68	518,61
16	26.05.2020	729,31	441,98	54	03.06.2021	673,14	537,35	92	14.06.2022	406,80	535,42
17	09.06.2020	501,88	530,26	55	15.06.2021	361,51	504,59	93	25.06.2022	261,46	515,47
18	21.06.2020	313,31	540,87	56	26.06.2021	256,84	519,02	94	06.07.2022	233,25	537,25
19	02.07.2020	237,60	533,88	57	07.07.2021	233,25	537,25	95	17.07.2022	256,84	519,02
20	13.07.2020	242,35	530,20	58	18.07.2021	261,46	515,47	96	29.07.2022	361,51	504,59
21	25.07.2020	328,07	529,68	59	31.07.2021	406,80	535,42	97	13.08.2022	624,37	508,09
22	08.08.2020	526,37	512,90	60	16.08.2021	701,68	518,61	98	27.08.2022	750,81	376,05
23	23.08.2020	753,71	428,29	61	29.08.2021	718,88	340,34	99	08.09.2022	717,82	314,22
24	04.09.2020	703,38	307,20	62	10.09.2021	721,98	321,10	100	20.09.2022	711,13	389,65
25	16.09.2020	721,75	355,50	63	22.09.2021	702,84	409,88	101	02.10.2022	636,43	537,30
26	29.09.2020	718,70	529,53	64	03.10.2021	571,92	512,49	102	11.10.2022	405,77	523,19
27	09.10.2020	470,43	548,90	65	12.10.2021	398,81	534,76	103	19.10.2022	312,05	548,19
28	17.10.2020	322,37	529,84	66	20.10.2021	307,25	556,97	104	26.10.2022	247,87	530,56
29	24.10.2020	252,94	518,75	67	27.10.2021	245,85	535,93	105	02.11.2022	241,51	559,75
30	31.10.2020	241,40	553,48	68	03.11.2021	242,19	562,23	106	09.11.2022	255,07	567,56
31	07.11.2020	249,08	567,62	69	10.11.2021	258,70	566,84	107	16.11.2022	289,10	553,12
32	14.11.2020	277,38	559,44	70	17.11.2021	295,52	549,33	108	23.11.2022	341,06	518,68
33	21.11.2020	324,65	530,31	71	24.11.2021	349,66	512,40	109	01.12.2022	469,11	531,92
34	29.11.2020	446,97	549,44	72	02.12.2021	480,35	522,90	110	10.12.2022	635,37	509,77
35	08.12.2020	610,41	530,71	73	12.12.2021	726,10	549,35	111	19.12.2022	735,60	423,96
36	17.12.2020	715,70	441,19	74	21.12.2021	753,72	408,21	112	27.12.2022	709,77	328,05
37	26.12.2020	789,98	376,53	75	29.12.2021	717,41	321,33				
38	03.01.2021	726,34	313,46	76	06.01.2022	724,54	315,05				

Самая первая покупка произошла из-за того, что закончился второй товар, когда первый товар был еще в наличии. Вторая покупка произошла из-за того, что закончился первый товар. Третья покупка произошла тоже из-за того, что закончился первый товар. Четвертая покупка произошла из-за того, что закончился второй товар. Пока мы будем считать, что нам известно, какая покупка чем вызвана, это нужно для следующего анализа (а чуть позже будем смотреть извне системы управления запасами, когда мы не будем знать эту информацию). Желтым (а также сиреневым) цветом выделены те покупки, которые вызваны событием завершения запаса.

Посмотрим на объем второй покупки для первого товара $y_2^1 = 783.85$. За какой период времени израсходовалась эта покупка? Объем этой второй покупки израсходовался за время от события совершения этой покупки t_2 до времени совершения следующей покупки t_3 , то есть $y_2^1 = \int_{t_2}^{t_3} f^1(t) dt$. Теперь взглянем на объем первой покупки второго товара $y_1^2 = 535.8$. Но мы не можем сказать, когда закончился объем покупки y_1^2 , так как следующая покупка была спровоцирована тем, что покупатель пошел пополнять

запасы первого товара, а не второго. Также мы не знаем, когда закончился объем второй покупки второго товара y_2^2 , так как третья покупка также была спровоцирована тем, что закончился первый товар. Однако из-за того, что четвертая покупка была вызвана тем, что закончился второй товар, мы можем сказать, что объем товара $y_1^2 + y_2^2 + y_3^2$ закончился за время, когда была совершена первая покупка t_1 , до времени, когда была совершена четвертая покупка t_4 , т.е. $y_1^2 + y_2^2 + y_3^2 = \int_{t_1}^{t_4} f^2(t) dt$. Действительно, в момент времени t_1 закончился второй товар, но в моменты времени t_2 и t_3 второй товар еще не закончился, он закончился в момент времени t_4 , при этом в моменты t_2 и t_3 происходило пополнение запаса второго товара, поэтому объем покупок y_2^2 и y_3^2 надо прибавить к первой покупке y_1^2 . Получается, что те покупки, которые пополняют запас, когда запас еще не закончился, мы объединяем вместе (как объем одной большой покупки, суммируем) и говорим, что этот большой объем закончился за время от начала самой первой вошедшей сюда покупки до времени, когда последняя вошедшая покупка закончилась (то есть до времени следующей за ней покупки). Так, надо объединить покупки первого товара с номера $i = 3$ до номера $i = 12$, второго товара с номера $i = 13$ до номера $i = 16$, затем опять первого товара с $i = 16$ до $i = 22$ и так далее. В таблице сиреневым цветом выделены последние покупки из последовательности покупок, которые вызваны тем, что запас соответствующего товара закончился, начиная с этой последней покупки надо начинать суммирование. Оранжевым цветом выделены покупки, объем которых заканчивается в момент возникновения следующей покупки. После отмеченных оранжевым цветом покупок начинается последовательность покупок, которые вызваны завершением запаса (отмеченные желтым цветом, затем последовательность опять завершается сиреневым цветом). Получается так, что происходит как бы переключение событий (завершения запаса) с первого на второй товар и наоборот (это происходит из-за того, что предпочтения между товарами меняются). Покупки, которые отмечены фиолетовым и оранжевым цветом, а также покупки между ними надо как раз объединять в одну большую покупку.

(Заметим, что объемы тех покупок, которые происходят тогда, когда соответствующий товар еще в наличии (не отмеченные желтым или сиреневым цветом), у нас получились как правило меньше (в нашей модели покупатель закупался до фиксированного максимума, хотя это может быть не так), так как покупателю нужно меньше этого товара. Объем тех покупок, которые произошли из-за завершения запаса соответствующего товара, приблизительно соответствует максимальному уровню запасов плюс некоторая величина на восстановление страховых запасов.)

В результате мы можем сформировать два набора данных $(t_i, y_i^{1'})$ и $(t_i, y_i^{2'})$, состоящие из объемов покупок первого и второго товара, причем в эти наборы также входят объединенные покупки, которые привязаны к дате совершения этой объединенной покупки (то есть к дате первой составляющей покупки из этого объединения покупок). Например, из данных покупок, содержащихся в таблице 1, можно получить два набора данных, представленных в таблице 2, где можно заметить, что объем третьей покупки первого товара $y_3^{1'} = \sum_{i=3}^{12} y_i^1 = 3878.73$, а объем первой покупки второго товара есть сумма $y_1^{2'} = y_1^2 + y_2^2 + y_3^2$ и так далее. Полученные два набора будут уже независимыми.

Таблица 2

ДВА НЕЗАВИСИМЫХ НАБОРА ДАННЫХ

Набор 1			Набор 2					
i	Дата t_i	Объем покупки y_i^1	i	Дата t_i	Объем покупки y_i^2	i	Дата t_i	Объем покупки y_i^2
1	06.01.2020	613,93	1	06.01.2020	1368,01	42	18.07.2021	515,47
2	15.01.2020	783,85	2	02.02.2020	541,25	43	31.07.2021	535,42
3	24.01.2020	3878,73	3	10.02.2020	557,89	44	16.08.2021	1589,94
4	16.04.2020	726,36	4	17.02.2020	535,62	45	03.10.2021	512,49
5	28.04.2020	723,11	5	24.02.2020	561,96	46	12.10.2021	534,76
6	11.05.2020	753,43	6	02.03.2020	566,96	47	20.10.2021	556,97
7	26.05.2020	2878,88	7	09.03.2020	549,70	48	27.10.2021	535,93
8	23.08.2020	753,71	8	16.03.2020	512,34	49	03.11.2021	562,23
9	04.09.2020	703,38	9	24.03.2020	520,32	50	10.11.2021	566,84
10	16.09.2020	721,75	10	03.04.2020	535,92	51	17.11.2021	549,33
11	29.09.2020	3914,32	11	16.04.2020	1638,52	52	24.11.2021	512,40
12	17.12.2020	715,70	12	09.06.2020	530,26	53	02.12.2021	522,90
13	26.12.2020	789,98	13	21.06.2020	540,87	54	12.12.2021	1981,72
14	03.01.2021	726,34	14	02.07.2020	533,88	55	25.01.2022	516,27
15	11.01.2021	709,77	15	13.07.2020	530,20	56	03.02.2022	562,28
16	20.01.2021	4568,95	16	25.07.2020	529,68	57	11.02.2022	574,03
17	23.04.2021	711,13	17	08.08.2020	1603,89	58	18.02.2022	545,14
18	05.05.2021	717,82	18	29.09.2020	529,53	59	25.02.2022	565,66
19	18.05.2021	2895,04	19	09.10.2020	548,90	60	04.03.2022	564,26
20	16.08.2021	701,68	20	17.10.2020	529,84	61	11.03.2022	540,92
21	29.08.2021	718,88	21	24.10.2020	518,75	62	19.03.2022	565,47
22	10.09.2021	721,98	22	31.10.2020	553,48	63	28.03.2022	546,23
23	22.09.2021	3853,09	23	07.11.2020	567,62	64	08.04.2022	1638,06
24	12.12.2021	726,10	24	14.11.2020	559,44	65	01.06.2022	518,61
25	21.12.2021	753,72	25	21.11.2020	530,31	66	14.06.2022	535,42
26	29.12.2021	717,41	26	29.11.2020	549,44	67	25.06.2022	515,47
27	06.01.2022	724,54	27	08.12.2020	2413,91	68	06.07.2022	537,25
28	15.01.2022	777,14	28	29.01.2021	509,77	69	17.07.2022	519,02
29	25.01.2022	4063,23	29	06.02.2021	531,92	70	29.07.2022	504,59
30	21.04.2022	758,33	30	13.02.2021	518,68	71	13.08.2022	1588,02
31	03.05.2022	721,98	31	20.02.2021	553,12	72	02.10.2022	537,30
32	16.05.2022	718,88	32	27.02.2021	567,56	73	11.10.2022	523,19
33	01.06.2022	2845,90	33	06.03.2021	559,75	74	19.10.2022	548,19
34	27.08.2022	750,81	34	13.03.2021	530,56	75	26.10.2022	530,56
35	08.09.2022	717,82	35	21.03.2021	548,19	76	02.11.2022	559,75
36	20.09.2022	4544,49	36	30.03.2021	523,19	77	09.11.2022	567,56
37	19.12.2022	735,60	37	11.04.2021	1587,97	78	16.11.2022	553,12
38	27.12.2022	709,77	38	03.06.2021	537,35	79	23.11.2022	518,68
			39	15.06.2021	504,59	80	01.12.2022	531,92
			40	26.06.2021	519,02	81	10.12.2022	1261,79
			41	07.07.2021	537,25			

Далее для каждого набора мы можем произвести восстановление функции скорости расхода продукции $\hat{f}^1(t) = \hat{\alpha}(t)\hat{f}(t)$ и $\hat{f}^2(t) = \hat{\beta}(t)\hat{f}(t)$, воспользовавшись методом восстановления функции по интегралам (так как у нас вновь получилось, что наблюдаемые значения расходуются за время от текущего события до следующего события, $y_i^1 = \int_{t_i}^{t_{i+1}} f^1(t) dt$ и $y_i^2 = \int_{t_i}^{t_{i+1}} f^2(t) dt$). По имеющимся данным из первого набора восстанавливаем функцию $\hat{f}^1(t)$ (рис. 3), а из второго набора функцию $\hat{f}^2(t)$ (рис. 4). На рисунках 3 и 4 горизонтальными линиями изображена средняя скорость расхода продукции $y_i/(t_{i+1} - t_i)$, которая выглядит как ступенчатая функция, причем ширина каждой ступеньки разная, где-то ступеньки идут часто, а где-то редко. Восстановление функции по интегралам заключается в том, чтобы подобрать такую функцию, чтобы площадь под графиком этой функции на каждом интервале приближалась к площади каждой ступеньки. Как было сказано выше, решение находится в виде интегрального кубического сглаживающего базисного сплайна со штрафной функцией. Обратите внимание, что данный метод совсем не похож на обычный метод сглаживания, если бы мы сглаживали ступенчатую функцию, никогда бы не получилось так, чтобы аппроксимирующая кривая опустилась ниже ступеньки и затем поднялась бы обратно на одном интервале (функция стремилась бы к самой этой ступенчатой линии, а у нас площадь под функцией стремится к площади под ступенькой).

Рис. 3. Восстановление из данных редких событий скорости расхода запасов первого товара $\hat{f}^1(t)$ (штук в день)

Рис. 4. Восстановление из данных редких событий скорости расхода запасов второго товара $\hat{f}^2(t)$ (штук в день)

После восстановления скорости расхода запаса продукции первого и второго товара $\hat{f}^1(t)$, $\hat{f}^2(t)$ получаем оценку исходной функции потребности $\hat{f}(t) = \hat{f}^1(t) + \hat{f}^2(t)$. Сравнение восстановленных функций с исходными функциями, заложенными при моделировании, представлено на рисунке 5. Затем определяем изменяющиеся предпочтения потребителя между первым и вторым товаром $\hat{\alpha}(t) = \hat{f}^1(t)/\hat{f}(t)$ и $\hat{\beta}(t) = \hat{f}^2(t)/\hat{f}(t)$. Сравнение восстановленных функций предпочтения с исходными представлено на рисунке 6. На обоих рисунках пунктирной линией указаны исходные, заложенные при моделировании функции.

Рис. 5. Сравнение восстановленных функций с исходными

Рис. 6. Сравнение восстановленных предпочтений с исходными

В результате получилось достаточно точно восстановить все исходные функции. Причем предположение о том, что потребитель придерживается своего постоянного уровня запасов, мы не использовали, потребитель мог каждый раз осуществлять покупки, не придерживаясь никакой определенной схемы. Однако в своих рассуждениях мы использовали то, что у нас заранее было определено, какая из покупок вызвана завершением запаса, а какая нет. Естественно, если мы будем анализировать данные покупок неподконтрольного нам потребителя, то такой информацией мы обладать не будем, у нас будут только даты и объемы купленной продукции одного и второго товара. Но оказывается, что из данных покупок потребителя можно распознать, какая покупка произошла вследствие завершения запаса, а какая нет, о чем будет рассказано во второй части статьи.

ЗАКЛЮЧЕНИЕ

В данной части исследования мы получили механизм, как из данных редких событий (т.е. покупок) восстанавливать динамику изменения предпочтений покупателя между двумя альтернативными товарами с помощью метода восстановления функции по интегралам. Были рассмотрены 4 схемы образования редких событий. Первая – потребитель совершает покупку только того товара, который закончился, и не покупает тот товар, который еще не закончился. При этой схеме поведения восстановление происходит элементарно, не требуется никакой дополнительной обработки данных, достаточно рассматривать покупки каждого товара отдельно. Вторая схема, когда запасы одного товара заканчиваются, потребитель начинает потреблять запасы другого товара до тех пор, пока другой товар не закончится, и только потом пополняет запасы обоих товаров. При второй схеме, к сожалению, не получится восстановить истинные предпочтения потребителя, нужно вводить дополнительные предположения, например, такие, что потребитель при покупке товара заранее знал, что товар не успеет ему надоесть (но получается так, что не изменение предпочтения влияет на покупки, а наоборот, покупки влияют на предпочтение). В третьей схеме покупки не привязаны к завершению запасов, а происходят по другим не связанным причинам. В этой схеме, если ввести предположение, что покупатель не изменяет свой максимальный уровень запасов со временем, то восстановление также происходит довольно просто, достаточно в исходных данных произвести сдвиг на одну пози-

цию (объем покупок y_{i+1} отнести к предыдущему моменту времени t_i). Наконец в четвертой схеме, когда потребитель пополняет запасы как первого, так и второго товара, если закончились запасы любого из товаров, также получается восстановить динамику изменения предпочтений потребителя. Однако в этом последнем случае анализировали данные исходя из того, что мы знаем покупка какого товара была вызвана завершением запаса, а какая нет, что на практике априори знать о неподконтрольном покупателе невозможно. В следующей части исследования будет показано, что на самом деле можно произвести распознавание событий, анализируя относительную среднюю скорость расхода запаса, и получить эту информацию.

ЛИТЕРАТУРА

1. Croston J.D. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* (1970-1977). 1972; 23(3):289-303.
2. Johnston F.R., Boylan J.E. Forecasting intermittent demand: a comparative evaluation of Croston's method. *Comment. International journal of forecasting*. 1996; 12(2):297-298.
3. Altman N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*. 1992; 46(3):175-185.
4. Cove T., Har P. Nearest neighbor pattern classification// *IEEE Transactions on Information Theory*. *IEEE Transactions on Information Theory*. 1967; 13(1), 21-27.
5. Walker S.H., Duncan D.B. Estimation of the probability of an event as a function of several independent variables. *Biometrika*. 1967. 54 (1/2): 167–178.
6. Efron B. and Tibshirani R.J. *An introduction of the Bootstrap*. New York: Chapman & Hall, 1993.
7. Willemain T.R., Park D.S., Kim Y.B., Shin K.I. Simulation output analysis using the threshold bootstrap. 2001;134(1):17-28.
8. Венцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения: учеб. пособие для вузов. 2-е изд., стер. М.: Высш. шк., 2000. 383 с.
9. Кораблев Ю.А. Емкостный метод определения функции скорости потребления // *Экономика и менеджмент систем управления*. Воронеж: Изд-во «Научная книга», 2015. №15(1.1). С. 140-150.
10. Кораблев Ю.А. Погрешность емкостного метода анализа редких событий, удаленность от конечного потребителя // *Известия Кабардино-Балкарского научного центра РАН*. Нальчик. 2019. № 3 (89). С. 48-77.
11. Кораблев Ю.А. Емкостный метод анализа редких событий в торговле различными товарами // *Бизнес. Образование. Право. Вестник Волгоградского института бизнеса*. 2019. № 3(48). С. 121-131.
12. Кораблев Ю.А. Разбор причин и оценка погрешности аномальных картин в емкостном методе анализа редких событий // *Экономика и управление: проблемы, решения*. М.: Научная библиотека, 2017. №8(6). С. 8-12.
13. Бауэрсокс Дональд Дж, Клосс Дэвид Дж. Логистика: интегрированная цепь поставок. 2-е изд. / Пер. с англ. Н.Н. Барышниковой, Б.С. Пинскера. М.: ЗАО «Олимп-Бизнес», 2008. 640 с.
14. Кораблев Ю.А. Метод восстановления функции по интегралам для анализа и прогнозирования редких событий в экономике // *Экономика и математические методы*. 2020. № 3 (в печати).
15. Barbera S., Hammond P.J., Christian S. *Handbook of utility theory*. Kluwer academic Publishers. 1998. Vol. 1.
16. Вершинина А.Г., Холодкова А.Е. Потребительские предпочтения как основа для формирования ассортиментной линейки мучных кондитерских изделий // *Азимут научных исследований: экономика и управление*. 2019. Т.8. № 3 (28). С. 105-108.

17. Мирная К.Ф., Ермолаева Е.О., Поздняковский В.М. Анализ потребительских предпочтений при выборе рулетов копчено-вареных из мяса птицы // Известия вузов, пищевая технология. 2015, № 1. С. 113-115.

18. Ng C.Y., Law K.M.Y. Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning // Computers and Industrial Engineering. 2020. Vol. 139.

REFERENCES

1. Croston J.D. Forecasting and stock control for intermittent demands. Operational Research Quarterly (1970-1977). 1972; 23(3):289-303.

2. Johnston F.R., Boylan J.E. Forecasting intermittent demand: a comparative evaluation of Croston's method. Comment. International journal of forecasting. 1996; 12(2):297-298.

3. Altman N. S. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician. 1992; 46(3):175-185.

4. Cover T., Hart P. Nearest neighbor pattern classification // IEEE Transactions on Information Theory. 1967; 13(1), 21-27.

5. Walker S.H., Duncan D.B. Estimation of the probability of an event as a function of several independent variables. Biometrika. 1967. 54 (1/2): 167-178.

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

7. Willemain T.R., Park D.S., Kim Y.B., Shin K.I. Simulation output analysis using the threshold bootstrap. 2001;134(1):17-28.

8. Ventzel E.S., Ovcharov L.A. *Teoriya sluchaynykh protsessov i yeye inzhenernyye prilozheniya: ucheb. posobiye dlya vtuzov. 2-ye izd., ster.* [The theory of random processes and its engineering applications: Textbook. allowance for technical colleges. 2nd ed.]. M.: Higher school., 2000. 383 p.

9. Korablev Yu.A. *Yemkostnyy metod opredeleniya funktsii skorosti potrebleniya // Ekonomika i menedzhment sistem upravleniya* [Capacitive method for determining the consumption rate function // Economics and management systems management]. Voronezh: Scientific Book Publishing House, 2015. №15 (1.1). Pp. 140-150.

10. Korablev Yu.A. *Pogreshnost' yemkostnogo metoda analiza redkikh sobytiy, udalennost' ot konechnogo potrebitelya* [The error of the capacitive method of analysis of rare events, remoteness from the end consumer] // News of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences. Nalchik, 2019. No 3 (89). Pp. 48-77.

11. Korablev Yu.A. *Yemkostnyy metod analiza redkikh sobytiy v trgovle razlichnymi tovarami* [A capacitive method for the analysis of rare events in the trade in various goods] // Business. Education. Law. Bulletin of the Volgograd Institute of Business. 2019. No. 3 (48). Pp. 121-131.

12. Korablev Yu.A. *Razbor prichin i otsenka pogreshnosti anomal'nykh kartin v yemkostnom metode analiza redkikh sobytiy // Ekonomika i upravleniye: problemy, resheniya* [Analysis of the causes and error estimation of anomalous pictures in the capacitive method of analysis of rare events // Economics and management: problems, solutions]. M.: Scientific library, 2017. № 8 (6). Pp. 8-12.

13. Bowersox Donald J., Kloss David J. *Logistika: integrirovannaya tsep' postavok. 2-ye izd.* [Logistics: Integrated Supply Chain. 2nd ed.] / Transl. from English N.N. Baryshnikova, B.S. Pinsker. M.: Olymp-Business CJSC, 2008. 640 p.

14. Korablev Yu.A. *Metod vosstanovleniya funktsii po integralam dlya analiza i prognozirovaniya redkikh sobytiy v ekonomike* [A method of recovering a function by integrals for analysis and forecasting of rare events in the economy] // Economics and Mathematical Methods. 2020. No. 3 (in press).

15. Barbera S., Hammond P.J., Christian S. Handbook of utility theory. Kluwer academic Publishers. 1998. Vol. 1.
16. Vershinina A.G., Holodkova A.E. *Potrebitel'skiye predpochteniya kak osnova dlya formirovaniya assortimentnoy lineyki muchnykh konditerskikh izdeliy* [Consumer preferences, as a basis for the formation of an assortment line of flour confectionery] // Azimuth of Scientific Research: Economics and Management. 2019.V. 8. No. 3 (28). Pp. 105-108.
17. Mirnaya K.F., Ermolaeva E.O., Pozdnyakovsky V.M. *Analiz potrebitel'skikh predpochteniy pri vybore ruletov kopcheno-varenykh iz myasa ptitsy* [Analysis of consumer preferences when choosing smoked-boiled poultry meat rolls] // University News, Food Technology. 2015. No. 1. Pp. 113-115.
18. Ng C.Y., Law K.M.Y. Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning//Computers and Industrial Engineering. 2020. Vol. 139.

MODELING AND RESTORATION OF CUSTOMER PREFERENCES BETWEEN TWO ALTERNATIVE PRODUCTS USING THE CAPACITY METHOD OF RARE EVENTS ANALYSIS IN THE ECONOMY* (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

This article proposes an approach to restore the dynamics of changes in consumer preferences between two alternative products from the rare events data associated with the purchases of these goods, based on the capacity method of rare events analysis. The main idea of the capacity method, in which rare events are analyzed from the point of view of the processes occurring inside the sources of events, is briefly described. Consumption processes, which are the most common in the economy during the events formation, can be modeled as processes of emptying / filling capacity. This allows to restore the change rate of this capacity level (the rate of products stock change) using the mathematical method of restoring the function by integrals. Four variants for the events formation with the replenishment of two alternative products stocks are considered. For each variant, suggestions on how to restore the changes dynamics in consumer preferences between two alternative products are presented. For the variant in which the consumer replenishes the stocks of both goods when any of the goods stock runs out, it is necessary to first perform the recognition task and determine whether the event is formed by end of stock of the first or second product. Recommendations for the implementation of this recognition are given. The restoration examples of the change dynamic in preference between two alternative products are shown.

Keywords: rare events, capacitive method, consumption rate, recovery, preference, alternative products.

Работа поступила 27.04.2020 г.

* The article was prepared based on the results of studies carried out at the expense of budgetary funds on the state assignment of the Financial University of 2020