

Л.В. Лабунец¹
Н.Л. Лебедева²
М.Ю. Чижов³

L.V. Labunets
N.L. Lebedeva
M.Yu. Chizhov

АНАЛИЗ ДИНАМИКИ ПОСТУПЛЕНИЙ В БЮДЖЕТ ОТ МОСКОВСКОГО РЫНКА НАРУЖНОЙ РЕКЛАМЫ

THE ANALYSIS OF THE DYNAMICS OF REVENUES FROM THE MOSCOW MARKET OF OUTDOOR ADVERTISING

В работе представлена методика анализа, моделирования и прогнозирования нестационарных временных рядов (НВР) поступлений в бюджет г. Москвы от объектов наружной рекламы. Показана возможность оценки основных статистик НВР по малому объему исходных данных. Проанализированы способы выбора параметров адаптивных прогнозных моделей НВР.

Ключевые слова: анализ временного ряда, локально взвешенная полиномиальная регрессия, робастное оценивание, адаптивное моделирование и прогнозирование временного ряда.

This paper presents a methodology for analysis, modeling and forecasting of nonstationary time series (NTS) revenues from Moscow of outdoor advertising. Shows the possibility of estimation for main statistics NTS in a small sample size. Analyzed the methods of selecting the parameters for adaptive predictive models of NTS.

Keywords: time series analysis, locally weighted polynomial regression, robust estimation, time series adaptive modeling and forecasting.

Введение

Эффективное управление различными сферами бизнеса предполагает накопление и оперативный анализ информации, выявление скрытых закономерностей и зависимостей, прогнозирование тенденций и рисков. Современная методология проектирования корпоративных систем в значительной мере опирается на методы, алгоритмические модели и средства хранения и интеллектуального анализа данных. Быстрый рост объемов информации, содержащей в скрытой форме знания, необходимые для формирования управленческих решений, привело к появлению современных программных средств, которые принято называть системами поддержки при-

нятия решений (СППР) [1]. Информационно-поисковые, оперативно-аналитические, а также интеллектуальные СППР обеспечивают, соответственно, возможность поиска необходимых данных, их группировку и обобщение, а также поиск функциональных и логических закономерностей в накопленных данных, построение моделей и правил, которые объясняют найденные закономерности или прогнозируют развитие некоторых процессов.

Подсистема хранения информации в СППР использует современные системы управления базами данных (СУБД) и концепцию хранилищ данных. Предложенная Эдгаром Коддом модель многомерных данных в виде так называемого информационного гиперкуба и сформулированные им правила формирования реляционных СУБД являются фундаментальной основой для проектирования указанной выше структурной компоненты СППР.

В данной работе исследуется многомерная модель данных в виде помесечных поступлений в бюджет г. Москвы от объектов наружной рекламы (ОНР) в период с 2008 по 2011 год.

¹ Доктор технических наук, старший научный сотрудник, зав. кафедрой информационных систем в экономике и управлении НОУ ВПО «Российский новый университет», профессор МГТУ им. Н.Э. Баумана.

² Главный специалист Департамента информационных технологий ОАО «Банк ВТБ».

³ Аспирант НОУ ВПО «Российский новый университет», зам. руководителя Департамента средств массовой информации и рекламы города Москвы.

Подсистема интеллектуального анализа данных использует современные методы и алгоритмы Data Mining. К типичным задачам этого раздела теории искусственного интеллекта относят такие, как формирование описательных и прогнозных моделей. Первые улучшают понимание человеком анализируемых данных с помощью поиска ассоциативных правил и алгоритмов кластеризации, т.е. группировки однородных объектов. Вторые призваны идентифицировать модели данных по набору обучающих пар «вход – выход» и выполнить прогнозирование результатов для новых наборов наблюдений. К такому виду относят задачи классификации объектов по их характеристикам и поиска регрессионных зависимостей.

В данной работе представлена методика анализа, моделирования и прогнозирования нестационарных временных рядов (НВР) помесечных поступлений в московский бюджет от ОНР по типам рекламы и городу в целом. Важной особенностью предлагаемой методики является ее адекватность экспериментальным измерениям в условиях малого объема выборки. Этот результат достигается за счет применения интеллектуальных моделей локально взвешенной полиномиальной регрессии и робастного оценивания основных статистик НВР.

Современная методология адекватного прогнозирования НВР основана на рациональном сочетании экспертных моделей данных, а также методов адаптивного анализа и моделирования временных рядов [2]. Экспертный подход к анализу исходной информации позволяет учесть ее неопределенность. Плодотворным в этом смысле является, на наш взгляд, применение систем, основанных на обработке знаний экспертов с помощью алгоритмов нечеткого логического вывода.

Важной особенностью адаптивного направления анализа НВР является учет ожиданий участников инвестиционного процесса, а также возможности возникновения и влияния случайных факторов. Исследование различных сценариев развития событий позволяет естественным образом учесть указанные выше особенности прогнозирования НВР. Существенное значение в этом смысле приобретает управление результатами прогноза с помощью параметров эконометрических моделей НВР. В данной работе представлен анализ различных сценариев прогноза поступлений в бюджет г. Москвы от ОНР на основе гибких адаптивных моделей динамики временных рядов.

Модель многомерных данных

Трехмерная модель исходных данных представляет собой информационный «куб» [1] со следующими измерениями:

1) «**время**» в виде последовательности месяцев $n = 1, 2, \dots, N$ ($N = 48$) в период с 2008 по 2011 год включительно;

2) «**округ**» в виде упорядоченной последовательности административных округов г. Москвы $m = 1, 2, \dots, M$ ($M = 11$), представленных в таблице 1 по состоянию на 2011 год;

3) «**тип ОНР**» в виде упорядоченной последовательности типов наружной рекламы $k = 1, 2, \dots, K$ ($K = 30$), представленных в таблице 2.

Меры, расположенные на пересечении осей измерений, представляют собой отчисления в московский бюджет от ОНР в миллионах рублей. СУБД поддерживает стандартные операции среза, вращения, консолидации и детализации над многомерной моделью данных [1].

Таблица 1

Административные округа

n_2	Наименование
1	Общегородской заказ (ОГЗ)
2	Восточный административный округ (ВАО)
3	Западный административный округ (ЗАО)
4	Зеленоградский административный округ (ЗелАО)
5	Северный административный округ (САО)
6	Северо-восточный административный округ (СВАО)
7	Северо-западный административный округ (СЗАО)
8	Центральный административный округ (ЦАО)
9	Южный административный округ (ЮАО)
10	Юго-восточный административный округ (ЮВАО)
11	Юго-западный административный округ (ЮЗАО)

В качестве примеров на рис. 1 и 2 представлены графики НВР отчислений в городской бюджет по месяцам. Рис. 1 иллюстрирует данные, полученные в результате консолидации по округам и среза для ОНР «Щит отдельно стоящий». Рис. 2 демонстрирует динамику поступлений, консолидированных как по округам, так и по типам рекламы.

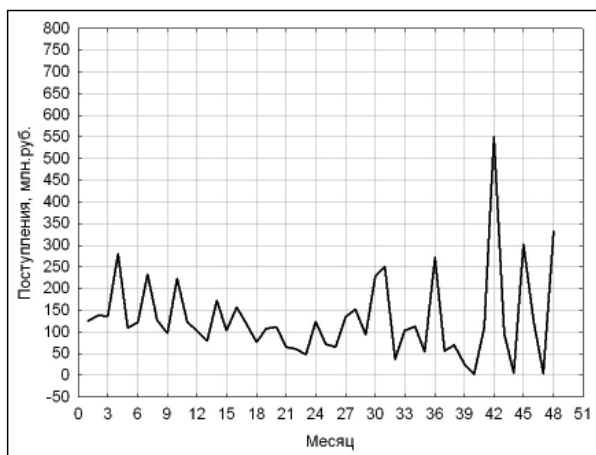


Рис. 1. Консолидированные по округам поступления от ОНР «Щит отдельно стоящий»

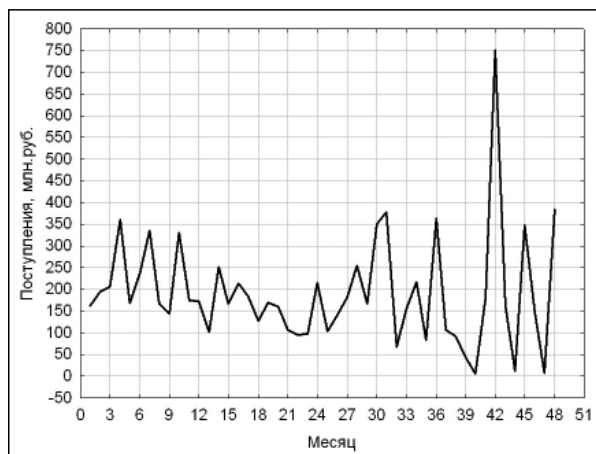


Рис. 2. Консолидированные поступления в бюджет от ОНР

Таблица 2

Типы наружной рекламы

n_3	Наименование
1	Витраж
2	Выносная щитовая конструкция (штендер)
3	Информационная конструкция предприятия и организации по обслуживанию населения на здании
4	Кронштейн на здании (площадью одной стороны более 1 кв. м и высотой более 1,5 м)
5	Кронштейн на здании (площадью одной стороны менее 1 кв. м и высотой менее 1,5 м)
6	Кронштейн на несветовой опоре
7	Кронштейн на световой опоре
8	Крышная установка
9	Маркиза

10	Настенное панно (площадью более 2 кв. м)
11	Настенное панно (площадью менее 2 кв. м)
12	Носимая реклама
13	Объемно-пространственная установка
14	Объемно-пространственный объект (отдельно стоящий)
15	Отдельно стоящая реклама на флагах и других мягких полотнищах
16	Прочие нетиповые объекты (мусорные контейнеры, телефонные будки и пр.)
17	Реклама на остановках транспорта
18	Реклама на строительной сетке
19	Реклама на флагах и др. мягких полотнищах
20	Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью более 2 кв. м)
21	Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью менее 2 кв. м)
22	Стационарная установка над проезжей частью
23	Транспарант-перетяжка
24	Тумба отдельно стоящая
25	Щит на временном ограждении
26	Щит на ограждении
27	Щит отдельно стоящий
28	Электронный экран на здании
29	Электронный экран на крыше
30	Электронный экран отдельно стоящий

В процессе анализа данных и поиска решений часто возникает необходимость в построении зависимостей между различными показателями. Число таких показателей может варьироваться в широких пределах. Таблицы реляционных баз данных не удовлетворяют требованиям объединения, просмотра и анализа информации с точки зрения множественности измерений. Оперативный многомерный анализ выполняют OLAP (On-Line Analytical Processing) – системы, предложенные Эдгаром Коддом в 1993 году. Основное назначение OLAP-технологии состоит в поддержке произвольных запросов аналитиков. Цель OLAP-анализа – проверка возникающих гипотез. Пример оперативного анализа структуры поступлений в московский бюджет от ОНР иллюстрирует рис. 3.

Визуальный анализ диаграмм рассеяния помесечных поступлений, сформированный OLAP-системой для 30-ти типов рекламы, позволяет аналитику ранжировать рекламные объекты по уровню их доходности. Вполне очевидной является следующая последовательность шести ОНР, упорядоченных по убыванию значимых суммарных поступлений за период с 2008 по 2011 год:

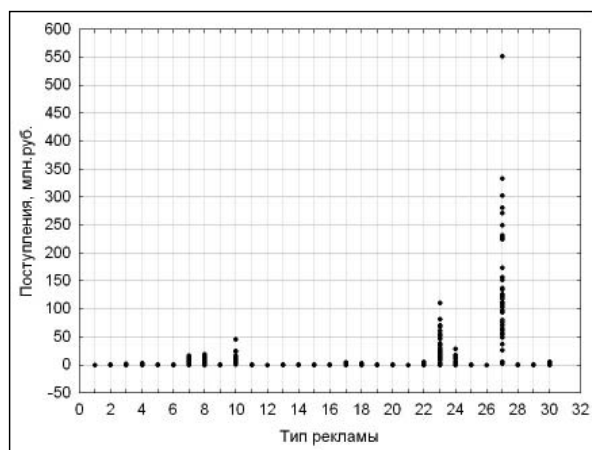


Рис. 3. Диаграммы рассеяния поступлений в бюджет от ОНР

- 1) щит отдельно стоящий (27) – 6301,21481116 млн руб.;
- 2) транспарант-перетяжка (23) – 1404,36650349 млн руб.;
- 3) настенное панно площадью более 2 кв. м (10) – 421,79276631 млн руб.;
- 4) крышная установка (8) – 321,92105990 млн руб.;
- 5) кронштейн на световой опоре (7) – 228,86775148 млн руб.;
- 6) тумба отдельно стоящая (24) – 177,54042206 млн руб.

Графики временных рядов поступлений в московский бюджет от этих типов рекламы представлены в логарифмическом масштабе на рис. 4. Поступления для первых трех типов (с номерами 27, 23 и 10 из таблицы 2) измерены по левой вертикальной шкале. Остальные – по правой шкале.

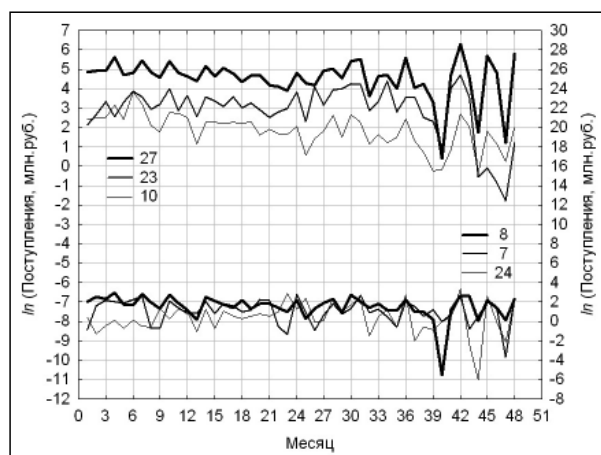


Рис. 4. Помесячные поступления от ОНР со значимой доходностью

Главная задача аналитика – это формирование гипотез на основе своих знаний и опыта. Гигабайты и терабайты накопленных данных содержат значительный объем «скрытых» знаний, которые человек не в состоянии исследовать самостоятельно. Существует высокая вероятность пропустить значимые гипотезы.

Методологию обнаружения «скрытых» знаний в больших объемах накопленной информации называют добычей данных (Data Mining). Data Mining – это алгоритмическое обнаружение в исходных данных «скрытых» знаний, которые ранее не были известны, нетривиальны, практически полезны и доступны для интерпретации человеком.

Рекуррентные статистики временного ряда

Рассмотрим одну из задач, решаемых с помощью подсистемы интеллектуального анализа данных СППР, а именно: задачу анализа и моделирования НВР $x(n)$, $n=1,2,\dots,N$ в виде консолидированных помесечных поступлений в бюджет г. Москвы от ОНР. При формировании эконометрической модели указанного временного ряда необходимо учитывать относительно малый объем выборки, содержащей $N = 48$ отсчетов, и наличие аномальных значений (см. рис. 2). Надежной методической основой для решения указанной выше задачи является применение рекуррентного сглаживания НВР [3] с помощью процедур локально взвешенной полиномиальной регрессии [4; 5]. В частности, для оценивания основных статистик НВР удобными, на наш взгляд, являются модели:

- ядерной робастной регрессии (Robust Locally Weighted Regression and Smoothing Scatterplots) [4];

- наименьших квадратов, взвешенных «расстоянием» (Distance Weighted Least Squares – DWLS) [5];

- экспоненциально взвешенной робастной регрессии (Exponential Weighted Regression – EWR) [6];

- комбинация моделей DWLS и EWR (DEWR) [3].

В алгоритме рекуррентного сглаживания параметрические модели начального ($i=1$) и остаточных ($i=2,3,\dots$) трендов $y_i(n-\tau)$ для отсчета относительного времени τ аппроксимируют локально постоянной $y_i(n-\tau) = c_0^{(i)}(\tau)$, линейной $y_i(n-\tau) = c_0^{(i)}(\tau) + \tau c_1^{(i)}(\tau)$ или квадратичной $y_i(n-\tau) = c_0^{(i)}(\tau) + \tau c_1^{(i)}(\tau) + \tau^2 c_2^{(i)}(\tau)$ зависимостью. Относительное время τ будем вычислять по схеме прогнозирования «вперед»,

т.е. $\tau = n - t$. Здесь t – временной отсчет в пределах интервала сглаживания $n - s_i N + 1 \leq t \leq n$, для которого оценивают значение начального или остаточного тренда, а $0,2 \leq s_i \leq 0,8$ – доля наблюдений, по которым выполняют сглаживание [4] на i -ой итерации удаления трендов [3]. В этом случае $0 \leq \tau \leq s_i N - 1$ и начало отсчета $\tau = 0$ соответствуют правой границе $t = n$ интервала сглаживания. Альтернативный выбор $\tau = n - t - (s_i N - 1)/2$ отвечает «центрированной» оценке тренда, поскольку в этом случае $|\tau| \leq (s_i N - 1)/2$ и начало отсчета $\tau = 0$ соответствуют середине $t = n - (s_i N - 1)/2$ интервала сглаживания.

В дальнейшем для определенности будем рассматривать квадратичную аппроксимацию, которая, в большинстве случаев, позволяет хорошо оценивать тренды, а также их темп изменения и кривизну, в особенности в окрестностях локальных экстремумов исходного и остаточного временных рядов.

Вектор $\bar{C}_i(\tau) = \{c_0^{(i)}(\tau), c_1^{(i)}(\tau), c_2^{(i)}(\tau)\}$ оптимальных коэффициентов полинома на i -ой итерации сглаживания выбирают из условия минимума

$$\bar{C}_i(\tau) \Big|_{opt} = \underset{\bar{C}_i(\tau)}{Arg \min} \left\{ \sum_{j=0}^{J_i-1} w_1^{(i)}(\tau - j) w_2^{(i)} \rightarrow \rightarrow \{ r_i(n - j) \} r_i^2(n - j) \right\} \quad (1)$$

взвешенного квадрата текущего остатка [4; 5]

$$r_i(n) = r_{i-1}(n) - y_i(n), \quad i = 1, 2, \dots$$

Здесь $J_i = s_i N$; $w_1^{(i)}(\tau)$ и $w_2^{(i)}\{r_i(n - \tau)\}$ – локальный и робастный веса остатка [7]. Исходный временной ряд $r_0(n) = x(n)$ инициализирует алгоритм рекуррентного оценивания. Рациональными критериями завершения итераций являются сходимость средней энергии текущих трендов к нулевому уровню

$$e_i = \sqrt{\frac{\sum_{n=1}^N w_2^{(i)}\{r_i(n)\} y_i^2(n)}{N \sum_{n=1}^N w_2^{(i)}\{r_i(n)\}}} \leq \varepsilon_1, \quad i = 1, 2, \dots$$

или незначительное изменение средней энергии остаточных трендов за одну итерацию

$$|e_{i+1} - e_i| \leq \varepsilon_2, \quad i = 2, 3, \dots,$$

где $\varepsilon_1 > 0$ и $\varepsilon_2 > 0$ – выбранные пользователем достаточно малые уровни значимости критериев. Критерий окончания итераций $i \leq I_{max}$ страхует неудачный выбор допусков. Финальные модели остаточного ряда и тренда рассчитывают по формулам

$$r(n) = r_i(n), \quad y(n) = x(n) - r(n),$$

где $I \leq I_{max}$ – номер заключительной итерации.

Локальный вес $w_1^{(i)}(\tau - j)$ текущего остатка $r_i(n - j)$ обратно пропорционален временному интервалу между отсчетами τ и j относительного времени. Поэтому значимый вклад в оценку параметров $\bar{C}_i(\tau)$ полиномиальной регрессии $y_i(n - \tau)$ вносят лишь те наблюдения $r_{i-1}(n - j)$, для которых отсчеты j близки к моменту времени τ по критерию веса $w_1^{(i)}(\tau - j)$.

С этой точки зрения масштаб s_i весовой функции удобно интерпретировать как параметр сглаживания данных. Вычислительные эксперименты показали, что для обеспечения быстрой сходимости процедуры рекуррентного оценивания в соответствии с указанными выше критериями текущий масштаб s_i рационально увеличивать по мере увеличения номера итерации, т.е.

$$s_i = \begin{cases} s_{i-1} + \Delta s, & s_{i-1} < S_{max} \\ S_{max}, & s_{i-1} = S_{max} \end{cases}, \quad i = 2, 3, \dots, I. \quad (2)$$

Локальный вес в модели Lowess $w_1^{(i)}(\tau) = Ker\{(\tau/s_i)/s_i\}$ определяется выбором положительной четной функции ядра $Ker(u)$, заданной на стандартном интервале $[-1; 1]$ и интегрируемой с единицей $\int Ker(u) du = 1$. Модели наиболее популярных ядер представлены в монографии [8, с. 140]. В частности, Клевеланд (Cleveland) рекомендует применять трижды кубическое ядро [4]

$$Ker(u) = \begin{cases} 70(1 - |u|^3)^3 / 81, & |u| \leq 1 \\ 0, & |u| > 1. \end{cases}$$

Начальный масштаб ядра s_1 выбирают с помощью процедуры скользящей проверки [8].

Локальный вес в модели DWLS удобно, на наш взгляд, вычислять с помощью формулы $w_1^{(i)}(\tau) = s_i N / \{v(s_i) \tau^2 + s_i N\}$, где $v(s)$ – скорость убывания веса в виде полинома, степень и коэффициенты которого выбирают эмпирически. В наших вычислительных экспериментах выбран полином пятой степени следующего вида

$$v(s) = 3,9301 - 33,9621s + 122,1792s^2 - 221,3187s^3 + 198,7453s^4 - 70,3125s^5.$$

График этого полинома представлен на рис. 5. В модели DWLS, как известно, отсутствует механизм робастного взвешивания остатков. Поэтому все веса $w_2^{(i)}\{r_i(n - j)\}$, $i = 1, 2, \dots, I$ полагают равными единице.

В алгоритмах, реализующих модели Lowess и DEWR, предусмотрен дополнительный (внут-

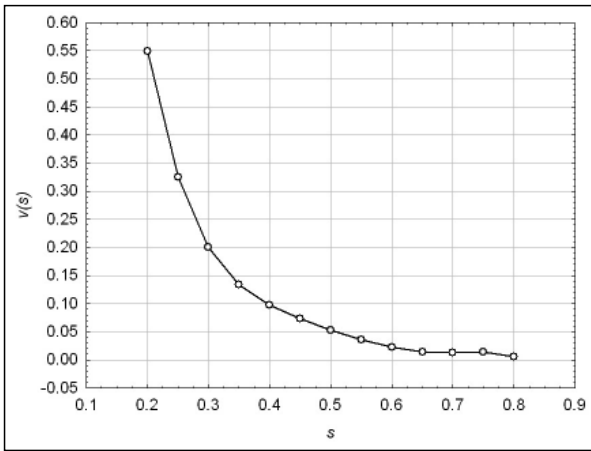


Рис. 5. Зависимость скорости убывания локального веса от параметра сглаживания модели DWLS

ренный) цикл робастного оценивания. Для фиксированной итерации i внешнего цикла локального взвешивания вес остатков $w_2^{(i)}\{r_i(n-j)\}$ на первой итерации робастного сглаживания полагают равным единице, поскольку отсутствует информация об остатке $r_i(n-j)$. На последующих итерациях вес обратно пропорционален остатку. Такая зависимость позволяет исключить влияние аномальных значений временного ряда на параметры $\bar{C}_i(\tau)$ полиномиальной регрессии $y_i(n-\tau)$ [6, 7]. Важно также отметить, что на второй и последующих итерациях робастного оценивания реализуют процедуру скользящего контроля, т.е. из суммы квадратов остатков в выражении (1) для целевой функции исключают отсчет $j=\tau$. Сглаживание продолжают до тех пор, пока робастные веса не стабилизируются. На практике процесс сходимости ограничивают двумя-тремя итерациями.

В модели Lowess Клевланд рекомендует применять робастное взвешивание остатков, предложенное Хубером (Huber). Весовая функция Хубера имеет вид

$$w_2^{(i)}\{r_i(n-j)\} = \begin{cases} r_i^2(n-j), & |r_i(n-j)| \leq R_i \\ 2R_i|r_i(n-j)| - R_i^2, & |r_i(n-j)| > R_i \end{cases}$$

где $R_i = 1,345RMS_i$; RMS_i – робастная мера масштаба остаточного НВР на i -ой итерации локально взвешенного сглаживания. В качестве такой меры рационально применять абсолютное медианное отклонение [9]. Различные варианты выбора робастных весовых функций представлены в работах [6; 7].

В комбинированной модели DEWR оптимальные параметры полиномиальной регрессии

выбирают из условия минимума функции потерь [3]

$$\bar{C}_i(\tau)|_{opt} = Arg \min_{\bar{C}_i(\tau)} \left\{ \sigma_i^{-\lambda/(1+\lambda)}(\tau) \sum_{j=0}^{J_i-1} w_1^{(i)} \rightarrow \rightarrow (\tau-j)w_2^{(i)}\{r_i(n-j)\} \right\}. \quad (3)$$

Здесь $w_2^{(i)}\{r_i(n-j)\} = \exp\{-\lambda r_i^2(n-j)/(2\sigma_i^2(n-j))\}$ – робастная весовая функция Л.Д. Мешалкина [6]; $\sigma_i^2(n-j)$ – экспоненциально взвешенная (ЭВ) оценка дисперсии остаточного НВР $r_i(n-j)$ на текущем временном интервале сглаживания $0 \leq j \leq s_i N - 1$; $\lambda \geq 0$ – параметр устойчивости ЭВ-регрессии к аномальным значениям данных.

Решение задач квадратичной оптимизации (1) и (3) приводит к системам линейных и нелинейных уравнений, соответственно [3]. Канонический вид этих систем позволяет решать их методом последовательных приближений. В качестве начальных оценок для тренда и дисперсии исходного и остаточных НВР выбирают стандартные оценки, полученные с помощью процедуры простой скользящей средней

$$y_i(n) = \frac{1}{J_i} \sum_{j=0}^{J_i} r_{i-1}(n-j),$$

$$\sigma_i^2(n) = \frac{1}{J_i} \sum_{j=0}^{J_i} r_i^2(n-j), \quad i=1,2,\dots,I.$$

Важным этапом формирования оценок основных статистик НВР является обоснование выбора параметров локально взвешенной полиномиальной регрессии.

Выбор параметров локально взвешенной полиномиальной регрессии

Как упоминалось выше, начальное значение параметра сглаживания s_1 обычно выбирают с помощью функции скользящей проверки (Cross-Validation). Альтернативный критерий оптимальности параметра s_1 , реализующий идею регуляризации модели локально взвешенной полиномиальной регрессии, основан на введении штрафа за ее сложность. В частности, рациональной, на наш взгляд, является целевая функция

$$s_1|_{opt} = Arg \min_{s_1} (DRF);$$

$$DRF = DR + \gamma DF, \quad (4)$$

где

$$DR = \frac{1}{N} \sum_{n=1}^N r^2(n); \quad DF = \frac{1}{N-2} \sum_{n=3}^N (\Delta^2 y(n))^2;$$

$\gamma > 0$ – вес штрафа за кривизну финального тренда

$$\Delta^2 y(n) = y(n) - 2y(n-1) + y(n-2),$$

$$n = 3, 4, \dots, N.$$

Уместно отметить, что целевая функция (4) аналогична критерию оптимальности цифрового фильтра Ходрика – Прескотта [3]. Кроме того, компоненту DF в выражении (4) удобно интерпретировать как дисперсию ошибки прогноза в среднем. Действительно, вторую разность $\Delta^2 y(n)$ финального тренда можно представить в виде ошибки

$$\Delta^2 y(n) = \delta(n) = y(n) - \tilde{y}_1(n-1)$$

наивного локально линейного прогноза

$$\tilde{y}_1(n-1) = y(n-1) + \delta_1(n-1)$$

тренда на один шаг времени. Здесь $\delta_1(n-1) = y(n-1) - y(n-2)$ – текущий темп изменения тренда. Иными словами, смысл критерия (4) состоит в выборе оптимального значения параметра s_1 модели локально взвешенной полиномиальной регрессии, обеспечивающего разумный компромисс между дисперсией DR ошибки $r(n)$ аппроксимации данных и дисперсией DF ошибки $\delta(n)$ локально линейного прогноза в среднем.

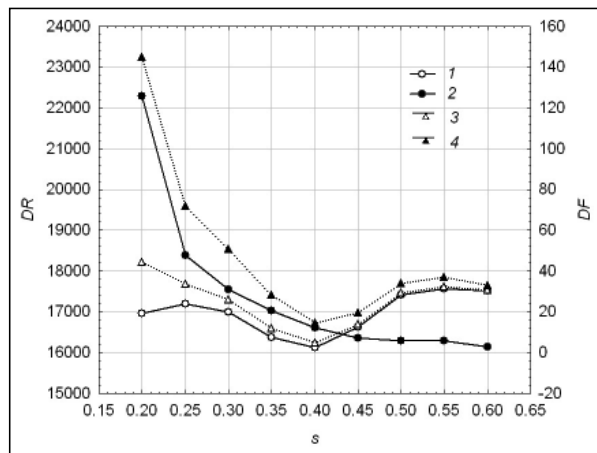


Рис. 6. Зависимость целевой функции от параметра сглаживания модели Lowess: 1 – дисперсия ошибки аппроксимации данных; 2 – дисперсия ошибки локально линейного прогноза тренда; 3 – целевая функция для $\gamma = 10$; 4 – целевая функция для $\gamma = 50$

Рис. 6 и 7 иллюстрируют зависимости целевой функции (4) от начального значения s_1 параметра сглаживания в рекуррентной процедуре оценки тренда консолидированных поступлений в московский бюджет от ОНП (см. рис. 2) для моделей Lowess и DWLS. Целевая функция, изме-

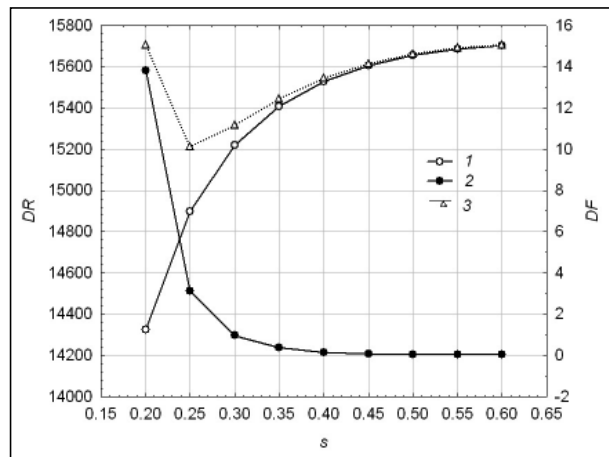


Рис. 7. Зависимость целевой функции от параметра сглаживания модели DWLS: 1 – дисперсия ошибки аппроксимации данных; 2 – дисперсия ошибки локально линейного прогноза тренда; 3 – целевая функция для $\gamma = 100$

ренная в масштабе левой вертикальной шкалы, соответствует различным значениям штрафа γ за ошибку прогноза.

В модели Lowess выполнялись три итерации робастного взвешивания. Параметры правила (2) управления локальным взвешиванием для указанных моделей сглаживания сведены в таблицы 3 и 4. Четвертый и пятый столбцы таблиц содержат значения средней энергии начального и финального остаточных трендов.

На рис. 8 представлены оценки трендов консолидированных поступлений в бюджет г. Москвы от ОНП, полученные с помощью моделей Lowess и DWLS с параметрами $s_1 = 0,4$ и $s_1 = 0,25$, соответственно.

Таблица 3

Параметры управления рекуррентным сглаживанием в модели Lowess для оценки тренда

s_1	S_{max}	I	e_1	e_i
0,2	0,6	10	26,184581	0,038549
0,25	0,65	13	25,781354	0,062346
0,3	0,7	15	25,887734	0,079519
0,35	0,75	10	25,694735	0,086252
0,4	0,8	17	25,185234	0,111986
0,425	0,8	11	25,066667	0,125209
0,45	0,8	14	24,984593	0,123597
0,5	0,8	14	24,749617	0,131539
0,55	0,8	14	24,716589	0,144613
0,6	0,8	10	24,722549	0,161205

Таблица 4

**Параметры управления рекуррентным
сглаживанием в модели DWLS
для оценки тренда**

s_1	S_{\max}	I	e_1	e_I
0,2	0,6	20	27,857751	0,004018
0,25	0,65	20	27,867429	0,004717
0,3	0,7	20	27,881724	0,005683
0,35	0,75	20	27,890537	0,006215
0,4	0,8	20	27,896649	0,004679
0,45	0,8	20	27,902536	0,003619
0,5	0,8	20	27,909339	0,002367
0,55	0,8	20	27,917260	0,001814
0,6	0,8	20	27,924725	0,002382

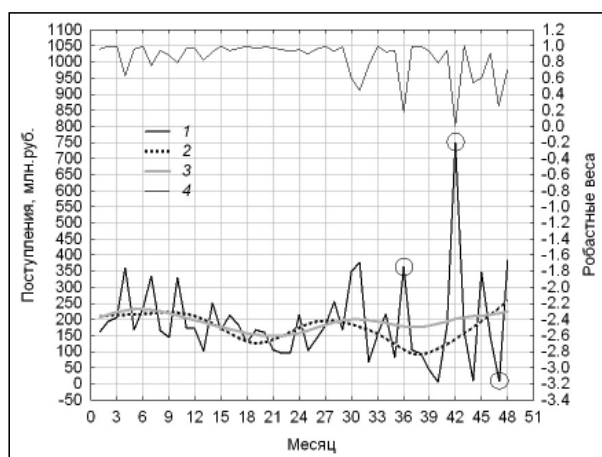


Рис. 8. Финальные тренды для различных моделей сглаживания:
1 – консолидированные поступления в бюджет от ОНП;
2 – модель Lowess; 3 – модель DWLS;
4 – робастные веса Хубера

Очевидно, что экспертные ожидания тенденции развития московского рынка наружной рекламы плохо согласуются с результатами, формируемыми моделью Lowess. Такой результат обусловлен весьма малыми величинами робастных весов Хубера, измеренных на рис. 8 по правой вертикальной шкале, для 36-го, 42-го и 47-го отсчетов НВР. Иными словами, вес штрафа за ошибку прогноза в среднем $10 \leq \gamma \leq 50$ дает несколько заниженное оптимальное значение $s_1 = 0,4$ параметра сглаживания для модели Lowess.

Достаточно хорошо согласованные оценки финального тренда обеспечивает выбор веса штрафа $\gamma = 100$ для модели Lowess и оптимальное значение $s_1 = 0,425$. Зависимость целевой функции (4) от параметра сглаживания s_1 для этого случая демонстрирует рис. 9.

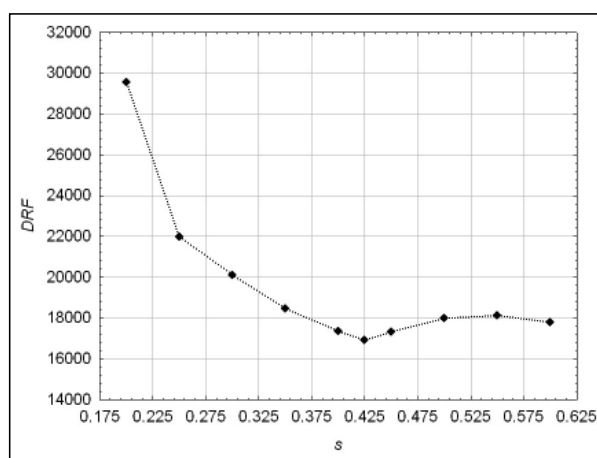


Рис. 9. Зависимость целевой функции от параметра сглаживания модели Lowess для $\gamma = 100$

Оценки трендов, соответствующие указанному выше выбору параметров моделей сглаживания Lowess и DWLS, представлены на рис. 10.

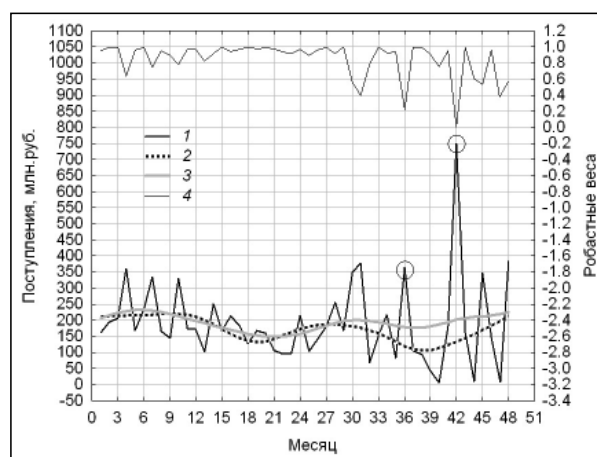


Рис. 10. Оптимальные оценки тренда для различных моделей сглаживания:
1 – консолидированные поступления в бюджет от ОНП;
2 – модель Lowess; 3 – модель DWLS;
4 – робастные веса Хубера

Рекуррентные робастные оценки тренда

Наряду с моделью Lowess на практике широкое распространение получила модель EWR – экспоненциально взвешенной робастной регрессии Л.Д. Мешалкина. Ее преимущества обусловлены, прежде всего, гибким механизмом управления робастным взвешиванием и высоким прогнозным потенциалом [6].

Анализ зависимости целевой функции (4) от параметра λ устойчивости статистик к аномальным значениям НВР показывает (рис. 11), что в широком диапазоне величин $10 \leq \gamma \leq 1600$

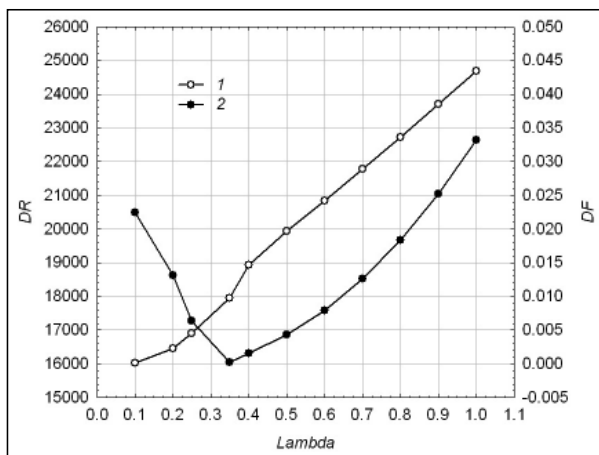


Рис. 11. Зависимость целевой функции от параметра λ устойчивости модели EWR: 1 – дисперсия ошибки аппроксимации данных; 2 – дисперсия ошибки локально линейного прогноза тренда

штраф $DRF(\lambda)$ практически не отличается от значений дисперсии $DR(\lambda)$ остаточного ряда. Поэтому для временного ряда консолидированных поступлений в московский бюджет от ОНП предпочтительным является величина $\lambda = 0,1$.

Процесс стабилизации робастных весов $w_2^{(i)} \{ r_i(n) \}$, $1 \leq n \leq N$ по критерию расстояния в N -мерном пространстве Евклида для этого случая иллюстрирует рис. 12.

Высокий потенциал модели экспоненциально взвешенной робастной регрессии для прогнозирования тренда НВР демонстрирует рис. 13. В зависимости от выбора значения λ темп изменения тренда за один шаг времени и робастные веса, измеренные в масштабе правой вертикальной шкалы, варьируются в широких диапазонах для правой границы сегмента данных.

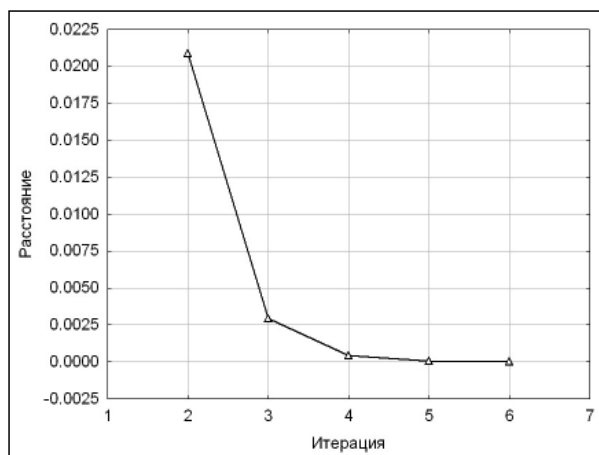


Рис. 12. Сходимость процедуры робастного взвешивания модели EWR для $\lambda = 0,1$

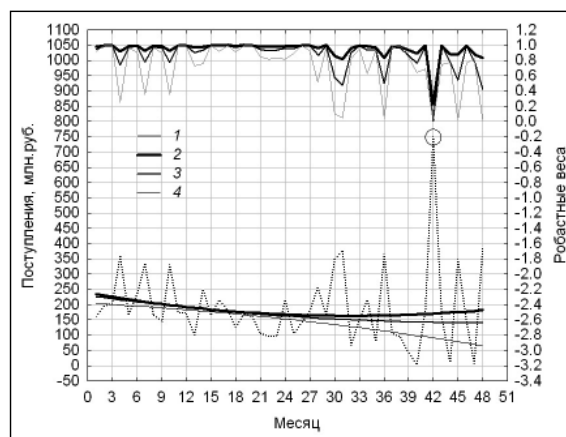
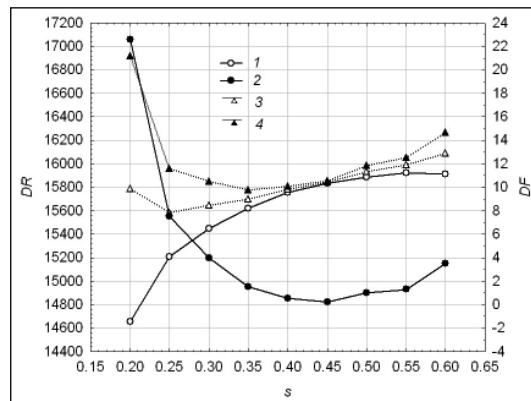
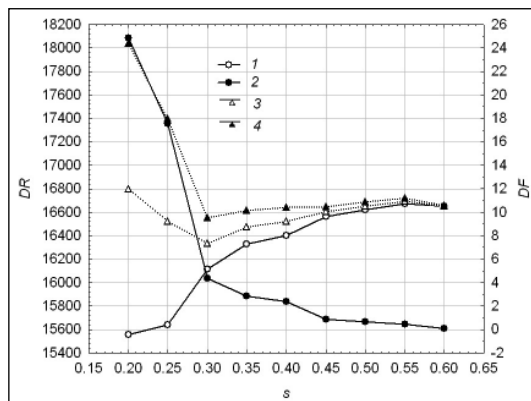


Рис. 13. Оценки тренда для модели сглаживания EWR: 1 – консолидированные поступления в бюджет от ОНП; 2 – тренд для $\lambda = 0,1$; 3 – тренд для $\lambda = 0,25$; 4 – тренд для $\lambda = 0,5$

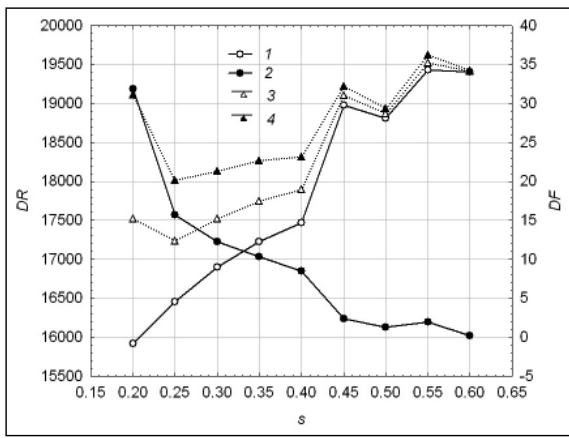
Иными словами, модель EWR обеспечивает возможность управления результатами прогноза в среднем с помощью адаптации параметра λ по мере поступления новых данных. Очевидно, что комбинация локального и робастного взвешивания, реализованная в модели DEWR [3], и соответственно адаптация по параметрам s_1 и λ может усилить указанное выше свойство.



а) $\lambda = 0,1$



б) $\lambda = 0,25$



в) $\lambda = 0,5$

Рис. 14. Зависимость целевой функции от параметра сглаживания модели DEWR:
 1 – дисперсия ошибки аппроксимации данных;
 2 – дисперсия ошибки локально линейного прогноза тренда;
 3 – целевая функция для $\gamma = 50$;
 4 – целевая функция для $\gamma = 100$

Рациональный выбор параметров s_1 и λ комбинированной модели сглаживания по критерию (4) обосновывает зависимости, представленные на рис. 14. Соответствующие оценки тренда консолидированных поступлений в бюджет г. Москвы от ОНР иллюстрирует рис. 15.

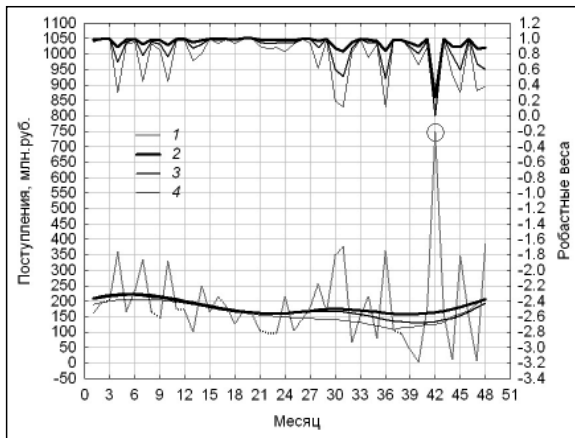


Рис. 15. Оценки тренда для модели DEWR с параметром сглаживания $s_1 = 0,35$:
 1 – консолидированные поступления в бюджет от ОНР;
 2 – тренд для $\lambda = 0,1$; 3 – тренд для $\lambda = 0,25$;
 4 – тренд для $\lambda = 0,5$

Наш выбор квазиоптимальных параметров модели DEWR состоит в следующем: $s_1 = 0,35$ и $\lambda = 0,1$. В этом случае результаты моделирования хорошо согласуются с экспертными ожиданиями темпа изменения тренда для московского рынка наружной рекламы. Более того, оценки остаточного НБР $r(n)$, $1 \leq n \leq N$ (рис. 16) для его правого сегмента практически не зависят

от выбора параметра устойчивости статистик к аномальным значениям в диапазоне значений $0,1 \leq \lambda \leq 0,5$.

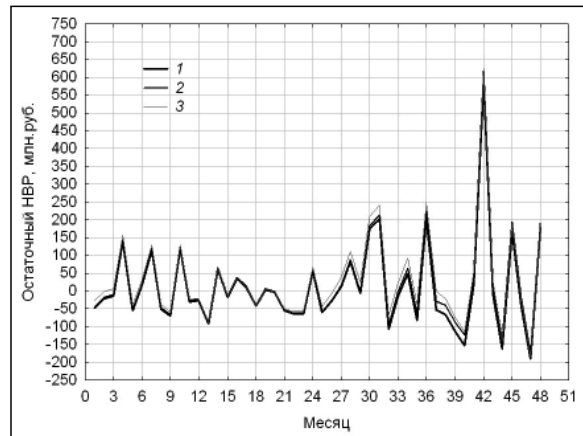


Рис. 16. Оценки остаточного НБР для модели DEWR с параметром сглаживания $s_1 = 0,35$:
 1 – $\lambda = 0,1$; 2 – $\lambda = 0,25$; 3 – $\lambda = 0,5$

Рекуррентные робастные оценки волатильности

Масштаб данных рационально оценивать с помощью алгоритма рекуррентного сглаживания квадрата остаточного НБР $r(n)$ [3]. В этой процедуре начальную ($i = 1$) и остаточные ($i = 2, 3, \dots$) меры волатильности $v_i^2(n)$ удобно аппроксимировать следующими локальными регрессионными моделями: $v_i^2(n - \tau) = g_0^{(i)}(\tau)$ – постоянной; $v_i^2(n - \tau) = g_0^{(i)}(\tau) + \tau g_1^{(i)}(\tau)$ – линейной; $v_i^2(n - \tau) = g_0^{(i)}(\tau) + \tau g_1^{(i)}(\tau) + \tau^2 g_2^{(i)}(\tau)$ – квадратичной. Здесь $0 \leq \tau \leq J_i - 1$ – относительное время, вычисленное по схеме прогнозирования «вперед». В дальнейшем для определенности будем рассматривать квадратичную аппроксимацию.

Вектор $\vec{G}_i(\tau) = \{g_0^{(i)}(\tau), g_1^{(i)}(\tau), g_2^{(i)}(\tau)\}$ оптимальных коэффициентов полинома на i -ой итерации сглаживания выбирают из условия минимума

$$\vec{G}_i(\tau) \Big|_{opt} = Arg \min_{\vec{G}_i(\tau)} \left\{ \sum_{j=0}^{J_i-1} w_1^{(i)}(\tau - j) w_2^{(i)} \rightarrow \right. \\ \left. \rightarrow \{ R_i^2(n - j) \} R_i^2(n - j) \right\}$$

взвешенного квадрата остатка

$$R_i(n) = u_{i-1}(n) - v_i^2(n), i = 1, 2, \dots$$

Здесь $w_1^{(i)}(\tau)$ и $w_2^{(i)} \{ R_i^2(n - \tau) \}$ – локальный и робастный веса. Квадрат остаточного НБР $u_0(n) = r^2(n)$ инициализирует алгоритм рекуррентного оценивания. На каждой итерации локального сглаживания выполняют нормировку $u_i(n) = u_{i-1}(n) / v_i^2(n)$, что обеспечивает схо-

димось текущей меры волатильности $v_i^2(n)$, $1 \leq n \leq N$ к единичному уровню в метрике Евклида. Соответственно стандартные критерии завершения итераций имеют вид:

$$E_i = \sqrt{\frac{\sum_{n=1}^N w_2^{(i)} \{ R_i^2(n) \} \{ v_i^2(n) - 1 \}^2}{N \sum_{n=1}^N w_2^{(i)} \{ R_i^2(n) \}}} \leq \delta_1, \quad (5)$$

$i = 1, 2,$

или $|E_{i+1} - E_i| \leq \delta_2, i = 2, 3, \dots,$

где $\delta_1 > 0$ и $\delta_2 > 0$ – выбранные пользователем достаточно малые уровни значимости критериев. В отличие от модели тренда, финальную оценку меры волатильности НВР рассчитывают по мультипликативной формуле

$$v(n) = \sqrt{\prod_{i=1}^I v_i^2(n)}.$$

Очевидно, что центрированная и нормированная структурные компоненты

$$z(n) = r(n) / v(n)$$

представляет собой временной ряд, стационарный по критериям тренда и волатильности.

Оптимальное начальное значение параметра сглаживания s_1 в рассмотренной выше рекуррентной процедуре оценки волатильности НВР будем выбирать в соответствии с критерием регуляризации, аналогичным (4). Компоненты штрафа DRF за сложность локально взвешенной полиномиальной регрессии в этом случае приобретают вид

$$DR = \frac{1}{N} \sum_{n=1}^N z^2(n); \quad DF = \frac{1}{N-2} \sum_{n=3}^N (\Delta^2 v(n))^2.$$

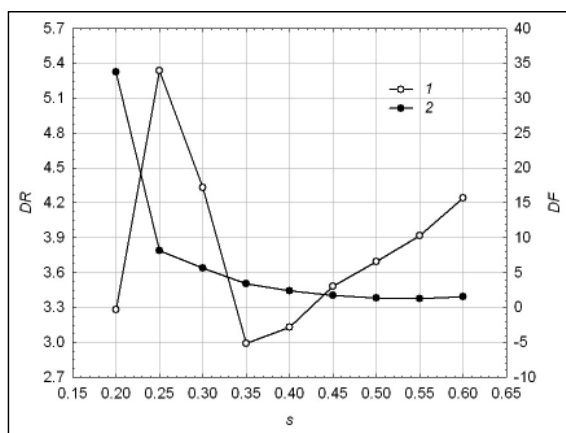


Рис. 17. Зависимости компонентов штрафа от параметра сглаживания модели Lowess: 1 – дисперсия ошибки аппроксимации квадрата остаточного НВР; 2 – дисперсия ошибки локально линейного прогноза меры волатильности

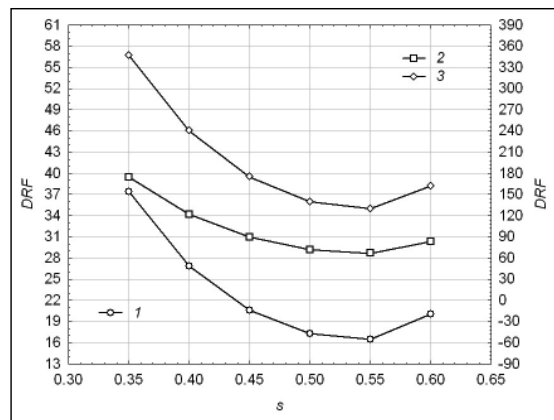


Рис. 18. Зависимости штрафа от параметра сглаживания в модели Lowess для оценки меры волатильности: 1 – $\gamma = 10$; 2 – $\gamma = 50$; 3 – $\gamma = 100$

Зависимости этих составляющих и целевой функции от параметра сглаживания s_1 для модели Lowess представлены на рис. 17 и 18, соответственно.

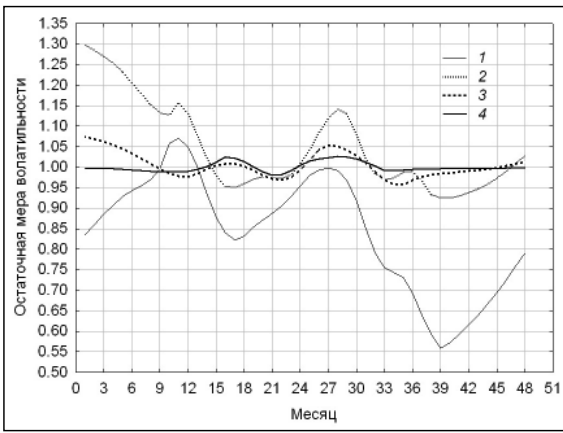
Штраф DRF на рис. 18 для значения веса $\gamma = 10$ измерен по левой вертикальной шкале, а для весов $\gamma = 50$ и 100 – по правой шкале. В вычислительном эксперименте выполнялись три итерации робастного взвешивания Хубера. Параметры правила (2) управления локальным сглаживанием сведены в таблицу 5.

Таблица 5

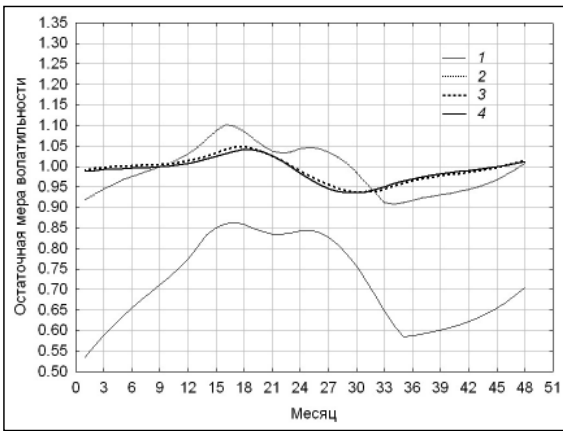
Параметры управления рекуррентным сглаживанием в модели Lowess для оценки волатильности

s_1	S_{max}	I	E_1	E_I
0,2	0,6	10	1367,6288	0,000953
0,25	0,65	13	1377,9393	0,000512
0,3	0,7	15	1377,4365	0,000570
0,35	0,75	10	1387,0383	0,000388
0,4	0,8	8	1375,1016	0,000521
0,45	0,8	22	1373,9722	0,001165
0,5	0,8	23	1370,7831	0,001605
0,55	0,8	20	1351,5053	0,001988
0,6	0,8	20	1323,5300	0,002489

Четвертый и пятый столбцы таблицы содержат значения критерия (5) для начальной $v_1^2(n)$ и финальной $v_I^2(n)$ остаточной меры волатильности. Процесс сходимости остаточных мер волатильности $v_i^2(n), 1 \leq n \leq N$ к единичному уровню в метрике Евклида по итерациям i алгоритма рекуррентного сглаживания демонстрирует рис. 19. Наибольшее абсолютное отклонение финального НВР $v_I^2(n), 1 \leq n \leq N$ от единицы не превышает величин 0,007182 и 0,029608 для значений параметра сглаживания $s_1 = 0,35$ и $0,55$.



а) $s_1 = 0,35$



а) $s_1 = 0,55$

Рис. 19. Сходимость остаточных мер волатильности к единичному уровню:

$$1 - v_2^2(n); 2 - v_3^2(n); 3 - v_5^2(n); 4 - v_7^2(n)$$

Финальные оценки меры волатильности $v(n)$, $1 \leq n \leq N$ консолидированных поступлений в московский бюджет от ОНР для указанных выше параметров модели сглаживания Lowess представлены на рис. 20. Робастные веса Хубера измерены по правой вертикальной шкале. Соответствующие оценки центрированного и нормированного временных рядов $z(n)$, $1 \leq n \leq N$ иллюстрирует рис. 21. Более предпочтительным, на наш взгляд, является выбор $s_1 = 0,425$ и $0,35$ при анализе тренда и масштаба данных.

Альтернативные DEWR-оценки меры волатильности консолидированных поступлений в бюджет г. Москвы от ОНР демонстрирует рис. 22. В вычислительном эксперименте выполнялось рекуррентное сглаживание квадрата центрированного ряда $r(n)$, $1 \leq n \leq N$, полученного с помощью модели DEWR с параметрами $s_1 = 0,35$ и $\lambda = 0,1$ (рис. 16). Предпочтительными, по нашему мнению, являются следующие параме-

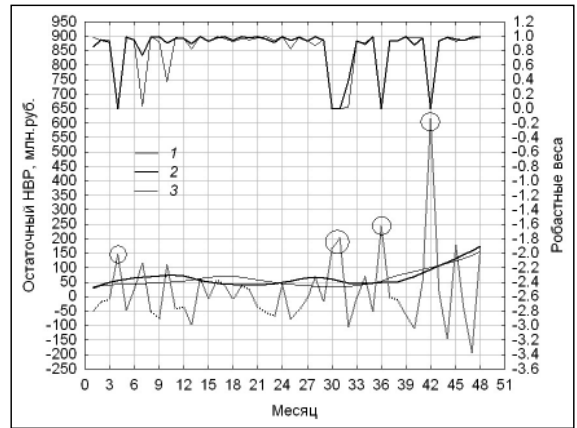


Рис. 20. Меры волатильности при сглаживании с помощью модели Lowess: 1 – $r(n)$ для $s_1 = 0,425$; 2 – $v(n)$ для $s_1 = 0,35$; 3 – $v(n)$ для $s_1 = 0,55$

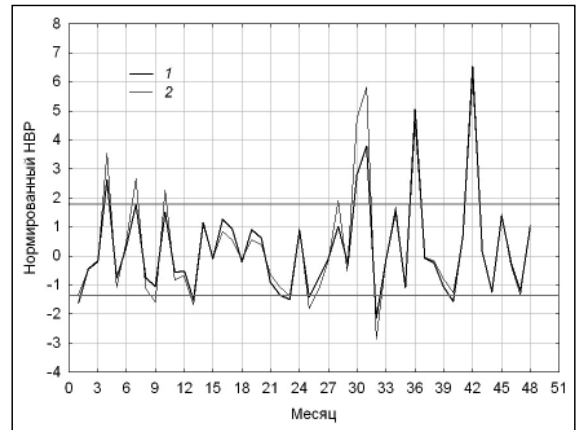


Рис. 21. Оценки центрированного и нормированных рядов при сглаживании с помощью модели Lowess: 1 – $s_1 = 0,35$; 2 – $s_1 = 0,55$

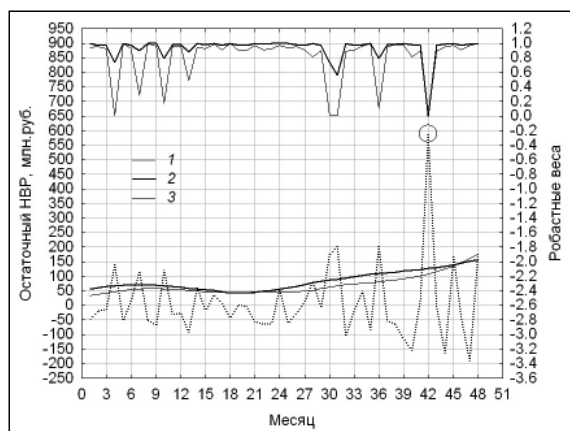


Рис. 22. Меры волатильности при сглаживании с помощью модели DEWR: 1 – $r(n)$ для $s_1 = 0,35$ и $\lambda = 0,1$; 2 – $v(n)$ для $\lambda = 0,1$; 3 – $v(n)$ для $\lambda = 0,25$

тры управления локальным сглаживанием ряда $r^2(n)$: $s_1 = 0,35$, $\Delta s = 0,05$; $S_{\max} = 0,75$, $I = 48$. Наибольшее количество итераций робастного взвешивания Л.Д. Мешалкина не превышало 10.

Сходимость по критерию (5) текущей меры волатильности $v_i^2(n)$, $1 \leq n \leq N$ к единичному уровню в метрике Евклида за двадцать итераций $2 \leq i \leq 20$ иллюстрирует рис. 23. Наибольшее абсолютное отклонение финального НБР $v_{48}^2(n)$, $1 \leq n \leq N$ от единицы не превышает величин 0,006273 и 0,000856 для значений параметра устойчивости статистик $\lambda_1 = 0,1$ и $0,25$. Соответствующие DEWR-оценки временного ряда $z(n)$, $1 \leq n \leq N$ представлены на рис. 24. В итоге приемлемый, на наш взгляд, выбор параметров DEWR модели сглаживания при анализе меры волатильности консолидированных поступлений в московский бюджет от ОНР состоит в следующем: $s_1 = 0,35$ и $\lambda = 0,1$.

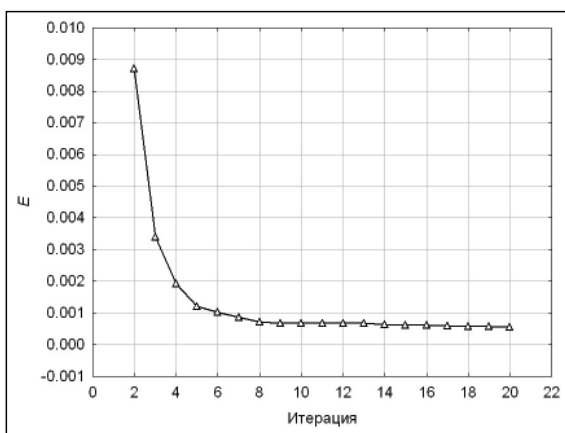


Рис. 23. Сходимость рекуррентной процедуры оценки волатильности НБР с помощью модели DEWR для $s_1 = 0,35$ и $\lambda = 0,1$

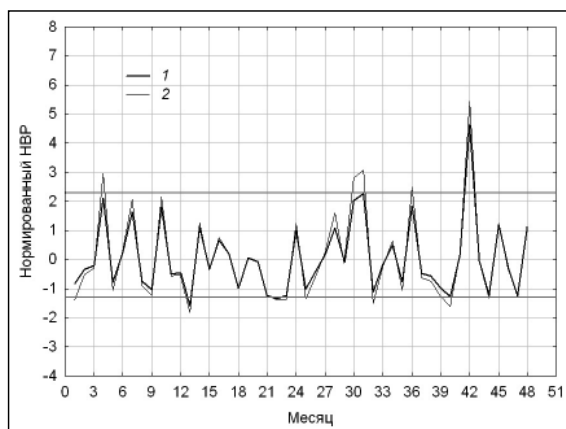
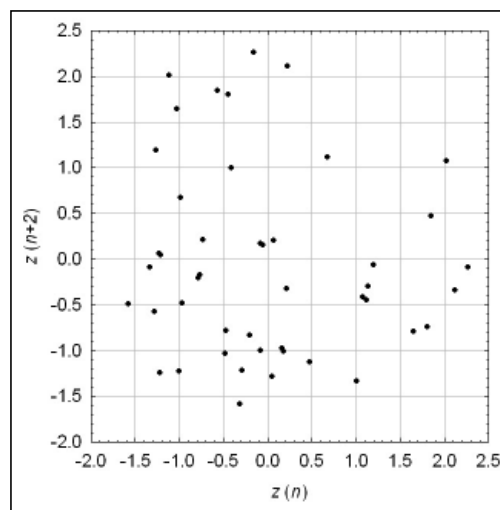


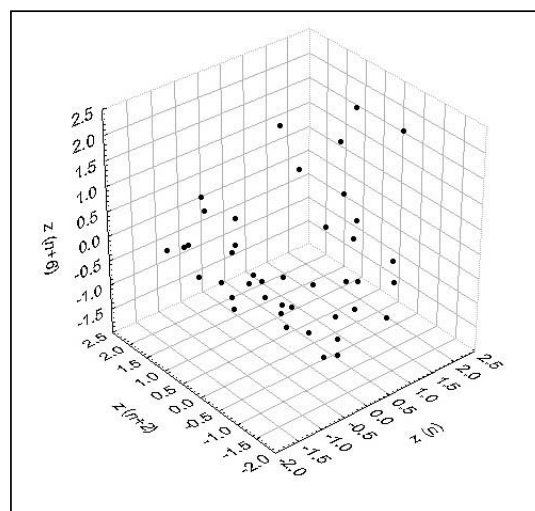
Рис. 24. Оценки централизованного и нормированного рядов $z(n)$ при сглаживании с помощью модели DEWR для $s_1 = 0,35$:
1 – $\lambda = 0,1$; 2 – $\lambda = 0,25$

Кластеризация доходности ОНР

Дальнейшее исследование динамики поступлений от рекламы связано с анализом статистик централизованного и нормированного временных рядов $z(n)$, $1 \leq n \leq N$. В частности, практический интерес представляет изучение корреляционных зависимостей этого ряда. Анализ формы диаграмм рассеяния для значений корреляционных лагов от 1 до 6 показал, что данные не обладают сколько-нибудь выраженным генеральным направлением. Указанное свойство иллюстрирует рис. 25, на котором представлены двумерная и трехмерная диаграммы для лагов двух и шести месяцев. Стандартные выборочные оценки коэффициента корреляции для этих лагов составили величины $-0,229 \pm 0,138$ и $0,207 \pm 0,132$, соответственно.



а)



б)

Рис. 25. Диаграммы рассеяния временного ряда $z(n)$: а) лаг два месяца; б) лаги два и шесть месяцев

Иными словами, отсчеты временного ряда $z(n)$, $1 \leq n \leq N$ практически не коррелированы. Обнаруженное свойство позволяет рассчитать доверительные интервалы и соответствующие им вероятности пребывания ряда в заданном диапазоне его значений. Удобной методической основой такого анализа является гистограмма, сглаженная сдвигом (Average Shifted Histogram – ASH) [8]

$$\varphi(z_k) = \frac{1}{N\delta} \sum_{j=1}^{m-1} w(j) \mathcal{G}_{k+j};$$

$$z_k = z_{\min} + k \Delta z; k = 0, \dots, K;$$

$$K = (z_{\max} - z_{\min}) / \Delta z.$$

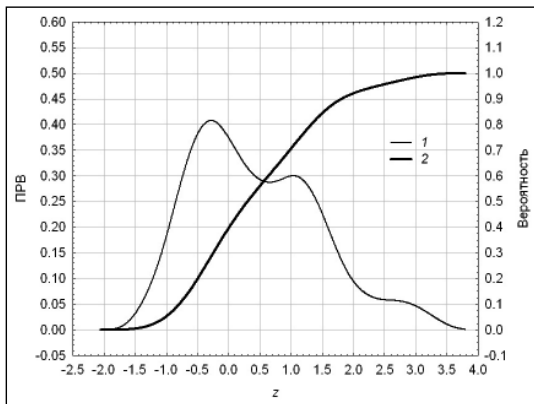
Здесь $\delta = 2IQ / \sqrt[3]{N}$ – робастная оценка ширины разрядных интервалов (bins) Фридмана – Дьякониса [8]; IQ – интерквартильный диапазон данных; $\Delta z = \delta / m$ и m – ширина «суженных» интервалов (narrow bins) и их количество; \mathcal{G}_l – количество наблюдений, попавших в l -ый «суженный» интервал ($\mathcal{G}_l = 0$, если $l < 0$ или $l \geq K$).

Окно данных $w(j)$ выбирают из условия $\sum_{j=1}^{m-1} w(j) = m$. В этом случае гистограмма интегрируема с единицей. Такой нормировке удовлетворяет обобщенное окно вида

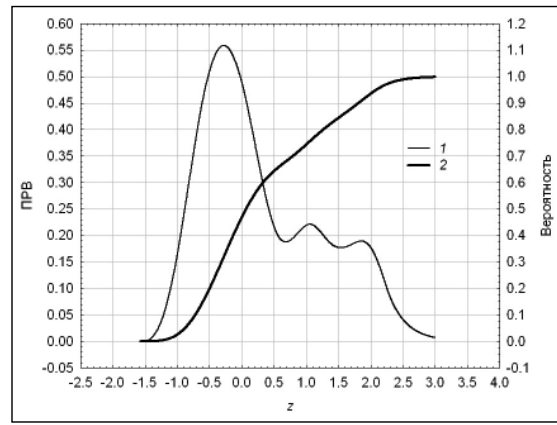
$$w(j) = m \text{Ker}(j/m) / \sum_{l=1}^{m-1} \text{Ker}(l/m).$$

Расчеты выполнялись для центрированных и нормированных временных рядов, полученных с помощью моделей Lowess (рис. 21) и DEWR (рис. 24) с параметрами $s_1 = 0,35$ и $\lambda = 0,1$. Пунктирными линиями на указанных рисунках отмечены уровни поддержки Z_S и сопротивления Z_R , выбранные по критерию робастных весов. В вычислительном эксперименте гистограммы сглаживались окном данных, сформированным на основе трижды взвешенного ядра Епанечникова [8]. Оценки распределений представлены на рис. 26.

Соответствующие вероятности $p_S =$



а) модель сглаживания Lowess



б) модель сглаживания DEWR

Рис. 26. ASH-оценки распределений отсчетов временного ряда $z(n)$:

1 – плотность распределения вероятностей (ПРВ);
2 – интегральное распределение

$= P\{z < Z_S\}$ и $p_R = P\{z > Z_R\}$ выбросов ряда $z(n)$, $1 \leq n \leq N$ за пределы уровней поддержки и сопротивления сведены в таблицу 6. Последний столбец таблицы содержит вероятности

$$p_{SR} = P\{Z_S \leq z \leq Z_R\} = 1 - p_S - p_R$$

для «типичных» значений доходности ОНР.

Таблица 6

Уровни поддержки и сопротивления централизованных и нормированных консолидированных поступлений от ОНР

Модель	Z_S	Z_R	p_S	p_R	p_{SR}
Lowess	-1,208046	2,231238	0,0233	0,059472	0,917228
DEWR	-1,214729	2,265476	0,0046	0,026147	0,969253

Важно также отметить, что ASH-оценки демонстрируют наличие кластерной структуры данных, а именно наличие трех классов «низкого» – *Low*, «среднего» – *Mid* и «высокого» – *High* уровней консолидированных поступлений от ОНР. Дискриминантные границы этих классов доходности ОНР, сформированные в виде абсцисс Z_{LM} и Z_{MH} локальных минимумов гистограмм, а также вероятности

$$p_L = P\{z < Z_{LM}\}, p_M = P\{Z_{LM} \leq z < Z_{MH}\},$$

$$p_H = P\{z > Z_{MH}\}$$

таких доходностей сведены в таблицу 7.

Таблица 7

Границы классов доходности ОНР

Модель	Z_{LM}	Z_{MH}	p_L	p_M	p_H
Lowess	0,653179	1,967250	0,604889	0,313794	0,081317
DEWR	0,679055	1,533509	0,679603	0,171313	0,149084

Рассмотренная выше методика интеллектуального анализа НВР позволяет формировать прогнозные модели доходности московского рынка наружной рекламы.

Адаптивная модель прогнозирования доходности ОНР

Будем предполагать, что изменения основных статистик для поступлений от ОНР г. Москвы достаточно медленные. Такого рода персистентность данных позволяет описывать тренд $y(n)$ и меру волатильности $v(n)$ для доходности ОНР с помощью адаптивной модели Хольта (Holt) [10]. В рамках этой модели локально линейные прогнозы

$$\hat{y}_1(n-1) = a_1(n-1) + a_2(n-1),$$

$$\hat{v}_1(n-1) = b_1(n-1) + b_2(n-1)$$

указанных выше статистик на один шаг времени рационально рассчитывать в терминах ошибок аппроксимации

$$\eta(n) = y(n) - \hat{y}_1(n-1),$$

$$\mu(n) = v(n) - \hat{v}_1(n-1), \quad 2 \leq n \leq N.$$

Адаптацию средних значений $a_1(n)$, $b_1(n)$ и темпов изменения $a_2(n)$, $b_2(n)$ статистик по мере поступления новых оценок $y(n)$ и $v(n)$ реализует прогнозная форма процедуры экспоненциального сглаживания

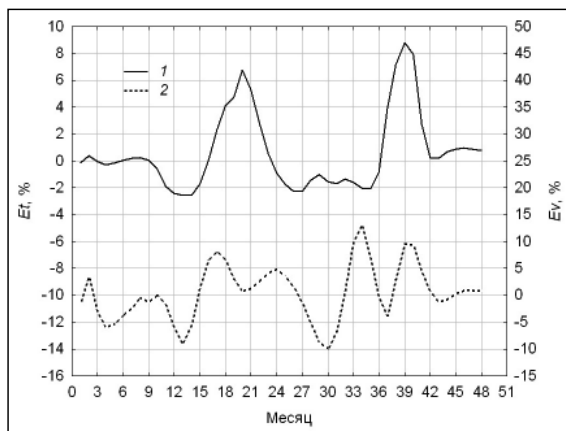
$$a_1(n) = a_1(n-1) + a_2(n-1) + \alpha_1 \eta(n),$$

$$a_2(n) = a_2(n-1) + \alpha_1 \alpha_2 \eta(n),$$

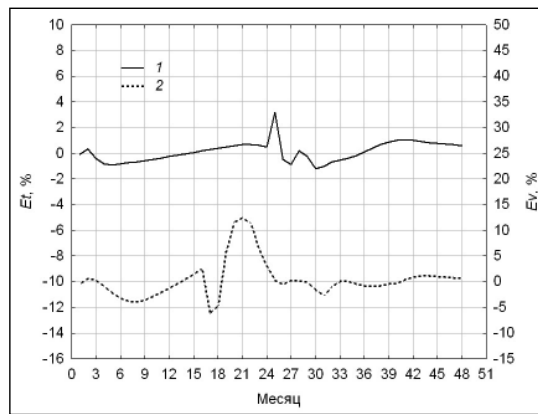
$$b_1(n) = b_1(n-1) + b_2(n-1) + \beta_1 \mu(n),$$

$$b_2(n) = b_2(n-1) + \beta_1 \beta_2 \mu(n).$$

Здесь α_1, β_1 и α_2, β_2 – параметры сглаживания статистик и их первых разностей. Эти параметры удовлетворяют известным ограничениям $0 < \alpha_1 < \alpha_2 \leq 1$ и $0 < \beta_1 < \beta_2 \leq 1$ [10]. Относи-



а) модель сглаживания Lowess



б) модель сглаживания DEWR

Рис. 27. Относительные ошибки аппроксимации статистик: 1 – тренда; 2 – меры волатильности

тельные ошибки аппроксимации

$$Et(n) = 100 \eta(n) / y(n),$$

$$Ev(n) = 100 \mu / v(n), \quad 2 \leq n \leq N$$

тренда и меры волатильности, полученные с помощью процедур рекуррентного сглаживания Lowess и DEWR, демонстрирует рис. 27.

Сравнение рисунков свидетельствует о том, что модель Хольта аппроксимирует DEWR-оценки статистик с меньшими ошибками. Соответствующие оптимальные значения параметров модели Хольта сведены в таблицы 8 и 9.

Таблица 8

Параметры модели Хольта для аппроксимации тренда

Модель	$a_1(1)$	$a_2(1)$	α_1	α_2
Lowess	210,1	1,025	0,736	1
DEWR	205,3	4,495	0,772	1

Таблица 9

Параметры модели Хольта для аппроксимации волатильности

Модель	$b_1(1)$	$b_2(1)$	β_1	β_2
Lowess	20,61	9,712	0,681	1
DEWR	53,81	3,173	0,396	1

Актуальные линейные прогнозы тренда и меры волатильности НВР на заданное количество m шагов времени приобретают вид

$$\hat{y}_m(N) = a_1(N) + m a_2(N),$$

$$\hat{v}_m(N) = b_1(N) + m b_2(N).$$

Соответствующие прогнозы

$$\hat{x}_m^S(N), \hat{x}_m^R(N), \hat{x}_m^{LM}(N), \hat{x}_m^{MH}(N)$$

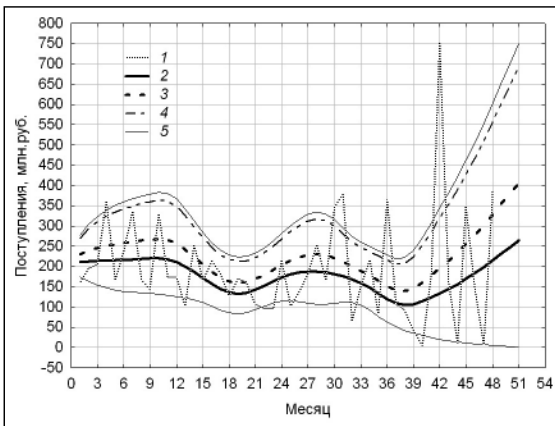
характерных уровней доходности ОНР вычисляются по формуле

$$\tilde{x}_m^{(*)}(N) = \hat{y}_m(N) + Z_{(*)} \hat{v}_m(N),$$

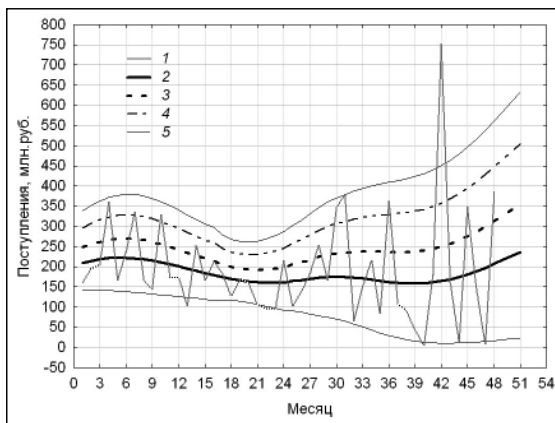
где под символом $(*)$ подразумевают один из четырех индексов S, R, LM, MH уровней доходности ОНР, представленных в таблицах 6 и 7.

Важно отметить, что каждому уровню $\tilde{x}_m^S(N)$, $\tilde{x}_m^R(N)$ или $\tilde{x}_m^{LM}(N)$, $\tilde{x}_m^{MH}(N)$ отвечают свои вероятности p_S, p_R, p_{SR} или p_L, p_M, p_H попадания доходности в соответствующие доверительные интервалы. В этом смысле прогнозные оценки уместно интерпретировать как интервальные. Результаты такого рода прогнозирования на три месяца ($m = 3$) поступлений в бюджет г. Москвы от ОНР с помощью моделей Lowess и DEWR иллюстрируют рис. 28.

Сравнение рисунков свидетельствует о том, что модель Lowess формирует несколько завышенные оценки тренда и меры волатильности для правого сегмента данных.



а) модель сглаживания Lowess



б) модель сглаживания DEWR

Рис. 28. Прогноз рынка наружной рекламы:
 1 – консолидированные поступления в бюджет; 2 – тренд;
 3 – граница «низкой/средней» доходности;
 4 – граница «средней/высокой» доходности;
 5 – линии поддержки и сопротивления

Заключение

В работе представлена методика интеллектуального анализа данных на примере исследования основных статистик и прогнозирования НВР в виде поступлений в бюджет г. Москвы от ОНР. Продемонстрирована возможность и эффективность оценивания тренда и меры волатильности рынка наружной рекламы с помощью моделей локально взвешенной робастной регрессии и процедуры рекуррентного сглаживания выборочных данных малого объема. Исходя из гипотезы медленного изменения статистик НВР, сформированы интервальные краткосрочные прогнозы динамики поступлений в рамках адаптивных эконометрических методов.

Литература

1. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2007. – 384 с.
2. Лукашин Ю.П., Рахлин Л.И. Современные направления статистического анализа взаимосвязей и зависимостей / отв. ред. Ю.П. Лукашин. – М. : ИМЭМО РАН, 2012. – 54 с.
3. Лабунец Л.В., Лабунец Н.Л., Чижов М.Ю. Рекуррентные статистики нестационарных временных рядов // Радиотехника и электроника. – 2011. – Т. 56. – № 12. – С. 1468–1489.
4. Cleveland W.S. Robust Locally Weighted Regression and Smoothing Scatterplots // Journal of the American Statistical Association. – 1979. – V. 74. – № 368. – P. 829–836.
5. McLain D.H. Drawing contours from arbitrary data points // The Computer Journal. – 1974. – V. 17. – № 4. – P. 318–324.
6. Шурыгин А.М. Прикладная стохастика: робастность, оценивание, прогноз. – М. : Финансы и статистика, 2000. – 224 с.
7. Воронцов К.В. Лекции по алгоритмам восстановления регрессии, 2009.
8. Scott D.W. Multivariate Density Estimation: Theory, Practice, and Visualization. – N.Y. : John Wiley & Sons, Inc., 1992. – 317 с.
9. Шестаков О.В. О скорости сходимости распределения выборочного абсолютного медианного отклонения к нормальному закону // Информатика и ее применения. – 2011. – Т. 5. – Вып. 5. – С. 74–79.
10. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов : учебное пособие. – М. : Финансы и статистика, 2003. – 416 с.