ОСОБЕННОСТИ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ ОПРЕДЕЛЕНИЯ ИХ ОПТИМАЛЬНЫХ БАЗОВЫХ ПАРАМЕТРОВ

Е.С. Кузнецов,
аспирант кафедры компьютерных технологий в проектировании и производстве Нижегородского государственного технического университета им. Р.Е. Алексеева

E-mail: yegor_s@rambler.ru
Адрес: г. Нижний Новгород, ул. Минина, д. 24

Рассмотрены особенности нового метода и алгоритмов прогнозирования, основанных на определении оптимальных базовых параметров процессов, оптимально дискретизированных по уровню и времени во временные ряды. Показано, что информационные системы прогнозирования, разработанные на основе этого метода, позволяют эффективно прогнозировать экономические процессы.

Ключевые слова: модель исходных данных, оптимальная дискретизация данных, оптимальные базовые параметры данных, прогнозирующие операторы, прогнозирование прогнозируемость.

1. Введение

Развитие прогностики как науки в последние десятилетия привело к созданию множества методов, процедур, приемов прогнозирования. По оценкам зарубежных и отечественных систематиков прогностики [1] насчитывается свыше ста методов прогнозирования. В связи с этим перед экономистами и специалистами других специальностей возникает задача выбора метода, который давал бы адекватные прогнозы для изучаемых систем и связанных с ними процессов.

Опыт показывает, что собственную оценку сложности реализации конкретного метода можно выполнить, если иметь чёткое математическое описание – математическую модель (ММ) конкретного метода, выраженную, например, в лингвистической форме в виде прогнозирующего оператора (ПО):

$$f(\{y_i\};\{p_j\}) \rightarrow \{y_{sf}\}\,, \qquad (1)$$

где $\{y_i\}, i \in [1, M]$ –исходный ряд данных длины M; $\{p_j\}, p_j \in [p_1, p_2, ..., p_n]$ – подбираемые параметры или оптимизируемые по выбранному критерию оптимальности; $\{y_{sf}\}, sf \in [1, 2, ..., L]$ – прогнозируемые выборки ряда; f – прогнозирующий оператор, характеризующий каждый из известных методов.

В практических исследованиях в качестве модели ПО, в основном, используются следующие функции f: линейная (ARMA, ARIMA) [2], квадратичная, степенная, показательная, экспоненциальная (экспоненциального сглаживания), логистическая. При этом не все процессы удаётся прогнозировать такими моделями, хотя в ряде случаев их удаётся заменять линейной комбинацией гармонических или иных функций.

В последнее время набирают популярность методы прогнозирования, ориентированные на обучение по прецедентам (относящиеся к разделу машинного обучения) или индуктивное обучение, основанные на выявлении общих закономерностей по частным эмпирическим данным. В этом случае, если линейная модель регрессии представляется необоснованной и предложить адекватную нелинейную модель $f(\{y_i\};\{p_j\}) \to \{y_{sf}\}$ также не удаётся, в качестве компромисса строится модель вида:

$$f(\{y_i\};\{p_j\}) = \sum_{j=1}^{n} \varphi_j f_j(\{y_i\}), \tag{2}$$

где φ_j – некоторые преобразования исходных признаков, в общем случае нелинейные. Задача состоит в том, чтобы подобрать неизвестные одномерные преобразования φ_j, при которых достигается минимум квадратичного функционала ошибок [3]. Поэтому чаще используются такие модели ПО, сложность идентификации параметров которых не сильно зависит от вида их нелинейностей.

Одной из главных задач в моделях прогнозирования является нахождение порядка n ПО, который определяется преимущественно числом её параметров [2] $\{p_j\}, p_j \in [p_1, p_2, ..., p_n]$ и определяет точность прогноза.

Заметим, что на практике иногда требуется прогнозировать непрерывные (аналоговые) процессы конечной длительности T. Во многих таких случаях частота дискретизации исходного сигнала f_d при дальнейших расчётах не меняется, что не соответствует строго теореме В.А. Котельникова [4-5]. Чаще она выбирается из соображений практики кратной секунде, минуте, часу и т.д., что свидетельствует о возможной потере необходимой информации в исходных данных уже на стадии дискретизации сигнала. Выбор f_d по Котельникову осложняется еще и тем, что не всегда просто указать верхнюю частоту f_d. В нашем случае, как увидим далее, модель ПО настраивается на оптимальную $f_d = 1/\Delta t$, учитывающую минимум потерь исходной информации.

2. Метод прогнозирования на основе оптимальных базовых параметров

В известных наиболее популярных методах определение (подбор) порядка модели и значений других параметров ПО, влияющих на точность прогнозирования модели, осуществляется независимыми между собой способами (только для n или только для Δt).

Поэтому нами выбрана такая модель ПО, при которой её параметры были бы согласованы между собой и находились (идентифицировались) по единому критерию специально введенных оптимальных базовых параметров (ОБП) [6-8].

Модель основана на предварительной дискретизации исходных векторных процессов продолжительности T одновременно по времени, с периодом $\Delta t = T/M$, и по значению в q-уровневые временные ряды исходной длины M

$$y = \left\{ \begin{matrix} y_0^1, ..., y_{n-1}^1, ..., y_{l-n+1}^1, ..., y_l^1, ..., y_{k-n+1}^1, ..., y_k^1, ..., y_{M-1}^1 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_0^r, ..., y_{n-1}^r, ..., y_{l-n+1}^r, ..., y_l^r, ..., y_{k-n+1}^r, ..., y_k^r, ..., y_{M-1}^r \end{matrix} \right\}, \tag{3}$$

имеющие ограничения по диапазону изменения параметров $MM - y_i$ и продолжительности:

$$-\infty < y_{min}^v \le y_i^v(k) \le y_{max}^{vj} < \infty, k \in [0, M-1], v \in [1, r], \tag{4}$$

где r-количество компонент векторного процесса.

Определение значений ОБП заключается в нахождении такой «тройки» $\{\Delta t_{opt}, q_{opt}, n_{opt}\}$ или «пары» БП $\{q_{opt}, n_{opt}\}$ (если есть доверие к дискретизации исходных данных и $\Delta t_{opt} = \Delta t$), при которой энтропия (3) по БП временного ряда $\{y^v_k\}$ будет минимальна:

$$E_{y;min} = min(log_2 N(\Delta t, q, n; y_k^v)) = \\ = log_2 N_{y,min} = n_{y;opt} \cdot log_2 q_{y;opt}, \tag{5}$$

где $$q \in [q_{min}, q_{max}], n \in [n_{min}, n_{max}] N_{y_i;min} = q_{y;opt}^{n_{y;opt}} \tag{6}$$

При этом ПО на основе локализованных ОБП $\{\Delta t_{opt}, q_{opt}, n_{opt}\}$ представим в виде:

$$f(\{y_i^r\}, \{\Delta t_{opt}, q_{opt}, n_{opt}\}) \longrightarrow \{y_{sf}^r\} \quad (7)$$

Если изначально дан дискретный процесс с фиксированным шагом Δt, то для прогнозирования определяются только q_{opt} и n_{opt}. Если же изначально имеется выборка из непрерывного процесса с неоптимальным и требующим уточнения шагом Δt, то по имеющимся M отсчетам сигнал методом сплайнов восстанавливается в «непрерывный». Затем образуются новые наборы выборок процессов с разными шагами дискретизации в интервале

$$T/M_{max} \leq \Delta t \leq T/M_{min},$$

где T – длительность исходного процесса.

Тот набор из упомянутых наборов БП будет иметь число компонент $M_{opt} = T/\Delta t_{opt}$, на котором параметры q, n дают среди всех наборов выборок процессов наименьшее значение энтропии (3). Так определяется вся тройка ОБП $\{\Delta t_{opt}, q_{opt}, n_{opt}\}$.

3. Построение прогнозирующего оператора при $k < M$

Предлагаемый метод прогнозирования, используя только динамические параметры исходных данных, позволяет построить по ряду (2) ПО для любого $k = n, n+1, \ldots, M-1$ в виде q-значной логической функции с ОБП – $\{\Delta t, q, n\}$.

$$(y^1_{k+1}, \ldots, y^r_{k+1})^T = \quad (8)$$
$$f((y^1_{k-n+1}, \ldots, y^r_{k-n+1})^T, (y^1_{k-n+2}, \ldots, y^r_{k-n+2})^T, \ldots, (y^l_k, \ldots, y^r_k)^T) \equiv f_k,$$

или эквивалентной таблице истинности (ТИ) [8].

Строки ТИ ПО строятся по всем идущим подряд n членам ряда отсчетов и следующего за ними отсчета, в качестве прогнозируемого ими значения. Величина n, является порядком математической модели ПО данных (2).

Порядок прогнозирующего оператора может определяться как минимальное n, при котором по одной и той же n-последовательности отсчетов прогнозируются одинаковые значения.

Когда числа n и q определяют ОБП, то ТИ оказывается непротиворечивой и ПО может допускать минимизацию в базисах функций q-значной алгебры логики.

4. Модификация прогнозирующего оператора прогнозирование при $k \geq M$

Для прогнозирования неизвестных выборок вне заданного ряда (3) при $k \geq M$ требуется модификации алгоритма (8).

Прогнозирование при $k \geq M$ заключается в пошаговом построении продолжения ТИ с $M-n+1$-ой по $M+sf$-ю строку, где $M+sf = 1, 2, \ldots, L$, а L – номер максимального шага прогнозирования или т.н. «прогнозного горизонта» для пополнения выборок данных (8), имеющихся в исходной ТИ.

Для определения y_{M+sf} используется последовательное сравнение $M-n+sf$-ой n-последовательности со всеми n-последовательностями, уже имеющимися в исходной таблице, рассматриваемыми как опорные («эталонные») по критерию «минимума расстояния» между ними [8].

$$y_{M+sf} = \arg \min_{y_k \in [y_{n+1}, y_{n+2}, \ldots y_{M+sf-1}]} d_k^{w_j^i}(s_f) \quad (9)$$

где

$$d_k^i(s_f) = \sum_{j \in [1,n]} \sum_{j \in [1,r]} w_j^i \, | y^l_{k+1-j} - y^l_{M+s_f-j} |^p, \quad (10)$$
$$k = n+1, n+2, \ldots M+sf-1$$

Таблица 1.

Таблица истинности

Аргументы прогнозирующего оператора				Прогноз
$(y^1_0, \ldots, y^r_0)^T$	$(y^1_1, \ldots, y^r_1)^T$		$(y^1_{n-1}, \ldots, y^r_{n-1})^T$	$(y^1_n, \ldots, y^r_n)^T$
$(y^1_1, \ldots, y^r_1)^T$	$(y^1_2, \ldots, y^r_2)^T$		$(y^1_n, \ldots, y^r_n)^T$	$(y^1_{n+1}, \ldots, y^r_{n+1})^T$
…				…
$(y^1_{k-n+1}, \ldots, y^r_{k-n+1})^T$	$(y^1_{k-n+2}, \ldots, y^r_{k-n+2})^T$		$(y^1_k, \ldots, y^r_k)^T$	$(y^1_{k+1}, \ldots, y^r_{k+1})^T$
…				…
$(y^1_{M-n-1}, \ldots, y^r_{M-n-1})^T$	$(y^1_{M-n}, \ldots, y^r_{M-n})^T$	…	$(y^1_{M-2}, \ldots, y^r_{M-2})^T$	$(y^1_{M-1}, \ldots, y^r_{M-1})^T$
$(y^1_{M-n}, \ldots, y^r_{M-n})^T$	$(y^1_{M-n+1}, \ldots, y^r_{M-n+1})^T$	…	$(y^1_{M-1}, \ldots, y^r_{M-1})^T$	-

В критерии близости (9) используются весовые функции индекса $j = 1, 2, ..., n$ с типом веса $i = \{c, l, e, h\}$:

$$w^{(c)}_j = 1,\ w^{(l)}_j = 1 + (1 - j)/n,\ w^{(e)}_j = e^{-j},\ w^{(h)}_j = j^{-1} \quad (11)$$

Возможна модификация формулы (7) с учетом классов эквивалентности, каждому из которых соответствует одинаковое прогнозируемое значение исходного ряда [9].

5. Связь оптимальных базовых параметров с предсказуемостью и восстанавливаемостью

Из введенных обозначений и алгоритма прогнозирования (8) для векторного временного ряда (3), удовлетворяющего условиям (4), существует ММ ПО, который по n_{opt} начальным, следующим подряд с шагом $\Delta t_{opt} \equiv T / M_{opt}$ q_{opt} -значным выборкам исходного ряда (3) позволяет вычислить все оставшиеся $M - n_{opt}$ выборок.

Исходя из этого, характеризуем «голографическое» свойство векторного M, q, n-процесса – «восстанавливаемость» (с точностью $1/q_{opt}$) с помощью прогнозирующего оператора (7) по n_{opt} известным, следующим подряд векторам-столбцам, последующих вектор-столбцов с любыми изменениями в них символов текста. И это позволяет естественным образом ввести понятие «предсказуемости (прогнозируемости)» (*Pr*) поведения векторного ПО, как отношения энтропии предсказываемой части векторных процессов к энтропии базовой (начальной части)

$$P_r = \left(\frac{\log_2 q_{opt}^{M-n_{opt}}}{\log_2 q_{opt}^{n_{opt}}} \right) + 1 = \frac{(M - n_{opt})}{n_{opt} + 1} = \frac{M}{n_{opt}} \le M, \quad (12)$$

из которого следуют следующие выводы:

Предсказуемость (12) векторного ряда (3) обладает следующими свойствами:

а) зависит явно только от длины M и порядка n_{opt} ПО (8);

б) не может превышать длины M векторного ряда;

в) возрастает с ростом M, если текст M, q, n при своем продолжении сохраняет ОБП, то есть при $\Delta M > 0$ остается стационарным «$M + \Delta M,\ q_{opt},\ n_{opt}$ -рядом»;

г) при $n_{opt} = 1$ предсказуемость *Pr* текста максимальна и равна M

Доказательство перечисленных свойств следует из возможности представления функции f ПО (5) в форме ТИ по тексту полной длины M [6].

6. Отличительные особенности предложенного метода

1. Предложен и проверен экспериментально на информационной системе прогнозирования метод, основанный на предварительной оптимальной дискретизизации («оптимальном загрублении») исходных данных во временные ряды.

2. Используемые ОБП $\{\Delta t, q, n\}$ находятся одновременно по единому энтропийному критерию, а не по различным известным ранее критериям оптимальности каждого из параметров, например, H.Akaike (только для n)[10] и В.А.Котельникова (только для $\{\Delta t\}$) [4].

3. Применяемый ПО един как для скалярных, так и для векторных процессов.

4. Метод позволяет уточнить частоту дискретизации исходных данных, если она была выбрана не оптимально.

5. Используемый ПО определяется в общем случае в виде нелинейной q-значной логической функции для любого прогнозируемого процесса.

6. Исходя из п.5 данный метод может быть применим в случае неизвестной функциональной зависимости прогнозируемого значения от предыдущих.

7. Прогнозирование экономических процессов

Экономические ряды имеют четкую восходящую или нисходящую тенденцию (тренд). Очевидно, что при этом количество уровней квантования будет не постоянным (не выполняется условие (4)). В данном случае целесообразно прогнозировать не сами значения ряда, а отклонения относительно предыдущего значения, т.е. разности 1-го порядка:

$$\overline{y^v_{\bar{k}}} = y^v_{\bar{k}+1} - y^v_{\bar{k}},\ \bar{k} \in [1, M-1], v \in [1, r] \quad (13)$$

Используя предположение, что количество уровней квантования для ряда (13) будет постоянным, после оптимальной дискретизации используя формулы (4)-(6) можно спрогнозировать ряд разностей 1-го порядка. Поскольку в конечном итоге важны не относительные изменения, а абсолютные значения, то используется обратное преобразование:

$$y^v_k = y^v_{\bar{0}} + \sum_{i=1}^{j-1} \overline{y^v_{\bar{i}}},\ j \in [1, M-1+L], v \in [1, r] \quad (14)$$

При прогнозировании рыночной стоимости акции также необходимо учесть объем сделок, который был совершен в этот же временной про-

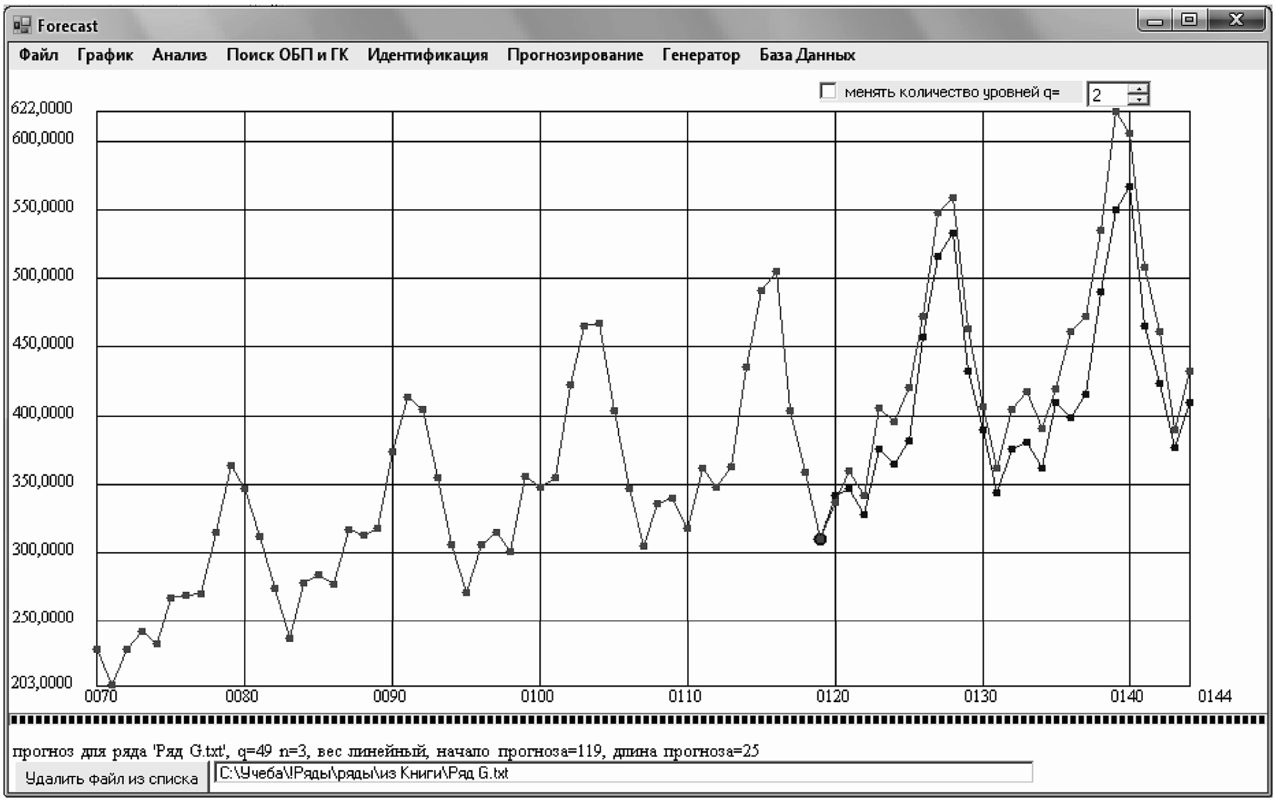

Рис. 1. Прогноз международных перевозок

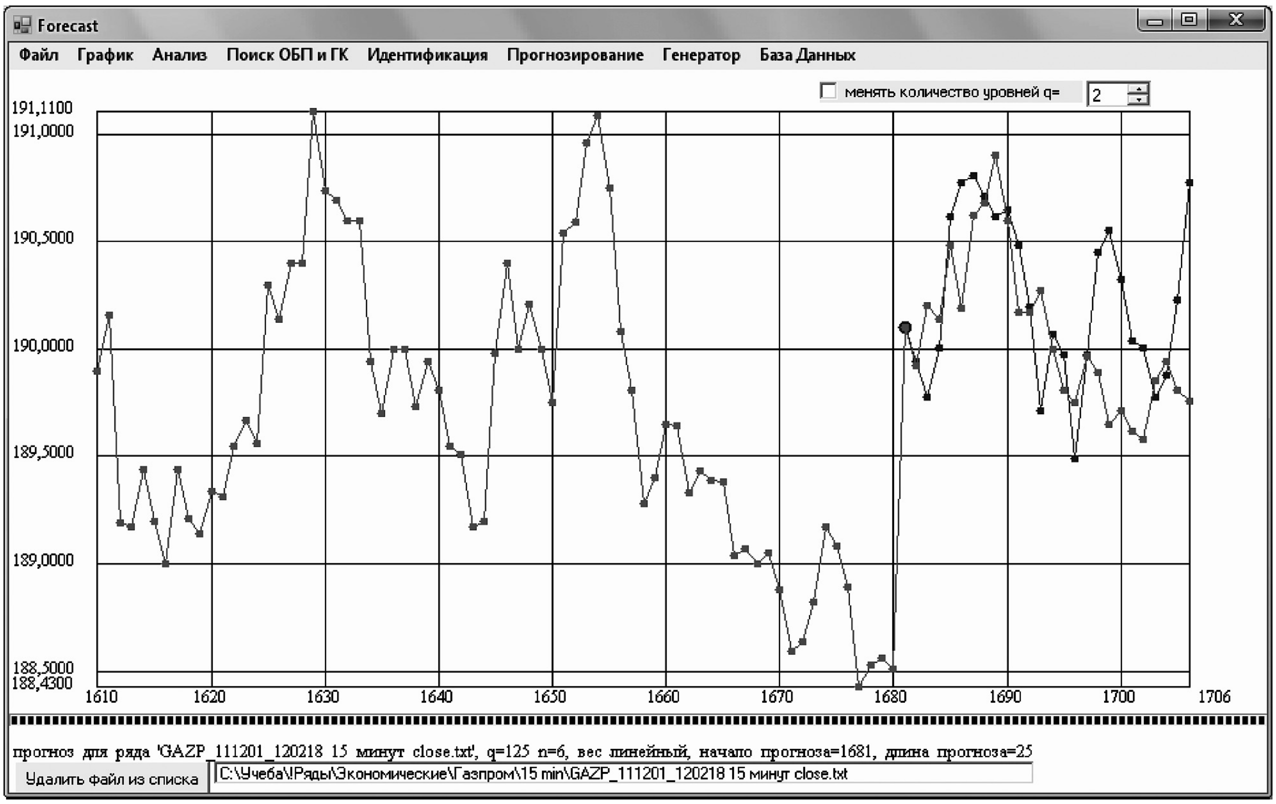

Рис. 2. Прогноз цены акции ОАО «Газпром»

межуток. Для этого прогнозный ряд можно представить в виде векторного процесса, состоящего из двух компонент: цена при окончании временного промежутка (Δt) и количество сделок, совершенных за данный временной промежуток времени.

8. Методика экспериментов. Оценка точности прогнозирования

Механизм верификации [1] прогноза позволяет выполнить оценку достоверности и точности прогноза по участку эталонного временного ряда. В качестве оценки точности прогноза используется формула

$$Er(L)=\sqrt{\frac{1}{L\text{-}1}\cdot\sum_{sf=1}^{L}\left|y_{M+sf\text{-}1}\text{-}y'_{sf}\right|^{2}}\ , \qquad (15)$$

где y' – исходный (реальный) процесс; s – шаг прогноза; L – длина прогноза. Сравниваются исходный процесс $(y'(s), s\in[M\text{-}L, M\text{-}1])$ и спрогнозированный процесс $(y(s), s\in[M\text{-}L, M\text{-}1])$ на основе исходного процесса с отброшенным концом $(y'(s), s\in[0, M\text{-}L\text{-}1])$. Поиск оптимальных базовых параметров и построение прогнозирующего оператора (12) осуществлялись на участке $k\in[0, M\text{-}L\text{-}1]$.

В качестве экспериментальных данных были выбраны 2 временных ряда: выборка цен акции ОАО «Газпром» за период *01.12.2011-18.02.2012* и «классический» ряд *Ряд G – Series G* [2], представляющий месячные международные авиаперевозки (в тысячах) в течение 12 лет с 1949 по 1960. Отличительной особенностью международных перевозок является наличие ярко выраженной сезонной составляющей. Ниже приведена таблица сравнения точности прогноза предложенного метода с двумя наиболее популярными методами – экспоненциального сглаживания и ARIMA.

Таблица 2.

Результаты прогнозирования международных авиаперевозок

$q_{opt}=49, n_{opt}=3;\ r=1,\ M=119;\ E=16{,}84;\ Pr=40$						
	Er(L)					
L	1	5	10	15	20	25
ОБП	3,65	19	24	26	34	34
Экспоненциальное сглаживание	17,5	14	24	23	30	35
ARIMA	7	23	33	37	53	57

Исходя из результатов, можно сделать вывод, что предложенный метод при прогнозировании временных рядов с ярко выраженной сезонной компонентой уступает методу прогнозирования, основанному на экспоненциальном сглаживании.

На *рис. 1* изображен результат прогнозирования методом на основе оптимальных базовых параметров, выполненный с помощью программы Forecast [11-12]. Синей линей изображен прогнозный ряд, красной – исходный ряд.

Для сравнения цен акции ОАО «Газпром» был спрогнозирован с учетом объемов сделок и без учета объема сделок. Результаты экспериментов (оценка точности прогноза рыночной стоимости акции и результаты предобработки) скалярного процесса, выполненных в программе Forecast [11-12], представлен в виде *табл. 3.*

Таблица 3.

Результаты прогнозирования скалярного процесса

$q_{opt}=125, n_{opt}=5, \Delta t_{opt}=15\,мин;\ r=1,\ M=1681;\ E=34{,}83;\ Pr=336$						
	Er(L)					
L	1	5	10	15	20	25
ОБП	0,73	0,48	3,6	4,4	4,3	4,3
Экспоненциальное сглаживание	0,96	1,6	2	2	2	2,1
ARIMA	0,48	0,7	0,6	0,75	0,95	1,05

На *рис. 2* изображен результат прогнозирования цены акции ОАО «Газпром» методом на основе оптимальных базовых параметров, выполненный с помощью программы Forecast [11-12]. Пунктирной линей изображен прогнозный ряд, сплошной – исходный ряд.

Результаты экспериментов (оценка точности прогноза рыночной стоимости акции и результаты предобработки) векторного процесса (с учетом объема сделок), выполненных в программе Forecast, представлен в виде *табл. 4.*

Таблица 4.

Результаты прогнозирования векторного процесса

$q_{opt}=106, n_{opt}=3, \Delta t_{opt}=15\,мин;\ r=2, M=1681;\ E=20{,}18;\ Pr=563$						
L	1	5	10	15	20	25
ОБП	0,27	0,46	0,55	0,5	0,7	1,1

9. Заключение

Таким образом, в статье предложен и проверен экспериментально новый метод прогнозирования экономических временных рядов, основанный на предварительной оптимальной дискретизизации («оптимальном загрублении») исходных данных во временные ряды. Данный метод зарекомендовал себя в задачах прогнозирования рядов, в которых априорная информация не позволяет сделать вывод о функциональной зависимости прогнозируемого значения от предыдущих. В тех случаях, когда имеется априорная информация, например, информация о наличии сезонной компоненты, следует использовать «классические», наиболее популярные методы прогнозирования, позволяющие учесть эту информацию.

Метод, основанный на применении ОБП, позволяет при прогнозировании рыночных курсов акций не только учесть объем сделок, но и одновременно спрогнозировать его.

Метод прогнозирования позволяет одновременно осуществлять и его верификацию (оценку достоверности и точности) по участку эталонного временного ряда.

Экспериментально было установлено, что при использовании информации об объеме сделок, возрастает прогнозируемость процесса и увеличивается в несколько раз точность прогноза, что в целом превышает точность прогнозирования ряда, выполненное «классическими» методами.

Предложенный алгоритм может применяться при решении практических задач бизнес-информатики, сводимых к задаче оптимальной дискретизации по уровню и по времени и к задаче прогнозирования временных рядов. ■

Литература

1. Прогностика. Технология. / Под ред. В.И.Сифорова. – М.: Наука, 1990.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974.
3. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. – М.: Наука, 1979.
4. Конева Е.С. Выбор моделей для реальный временных рядов // Автоматика и телемеханика. – №6. – 1988. – С. 3-18.
5. Романюк Ю.А. Основы цифровой обработки сигналов. В 3-х ч. Ч.1. Свойства и преобразования дискретных сигналов: Учебное пособие. – М.: МФТИ, 2005.
6. Кирьянов К.Г. Выбор оптимальных базовых параметров источников экспериментальных данных при их идентификации // Идентификация систем и задачи управления: тр. 3-й междунар. конф.– М.: ИПУ РАН, 2004. – С. 187-208.
7. Кирьянов К.Г. Идентификация динамических и информационных характеристик многоканальных систем на основе оптимальной дискретизации данных // Идентификация систем и задачи управления: тр. 9-й междунар. конф.– М.: ИПУ РАН, 2012. – С. 252-265.
8. Кирьянов К.Г., Кузнецов Е.С. Особенности прогнозирования дискретных и аналоговых векторных процессов на основе идентификации их базовых параметров // Труды 14-й Научной конференции по радиофизике.– Н.Новгород: ННГУ, 2010. – С. 278-279.
9. Кирьянов К.Г., Кузнецов Е.С. Модификация метода прогнозирования аналоговых и дискретных процессов в программе forecast 2 // Труды 12-й Научной конференции по радиофизике. – Н.Новгород: ННГУ, 2008. – С. 271-273.
10. Akaike H. A New Look at the Statistical Model Identification. // IEEE Transactions on Automatic Control. – 1974. – Vol. 19. – P. 716-723.
11. Кирьянов К.Г., Кузнецов Е.С. Информационная система прогнозирования векторных временных рядов // Информационные системы и технологии (ИСТ-2010): тез. докл. междунар. науч.-техн. конф.– Н.Новгород: НГТУ, 2010. – С.158-159.
12. Свидетельство о государственной регистрации программы для ЭВМ №2008611799, 09.04.2008