

Б.Ф.Кириянов, Ю.Ю.Петрова

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С «ОСОБЫМИ» ЗНАЧЕНИЯМИ

The method of forecasting of time rows with some values that are essentially different from typical. This method is used in addition to the classical method of sliding hole and it helps to make the forecasts much more precise when there are some particular values in the row.

Введение

Прогнозирование временных рядов, представляющих собой последовательности значений реализаций соответствующих величин, используется в цифровых системах управления различного назначения [1,2], экономике [3-5], сельском хозяйстве [6], медицине [7,8] и в ряде других областей и систем. Вопросы теории их прогнозирования рассматривались, например, в [2,5,9-13]. В частности, в [11] рекомендуется учитывать при прогнозировании особенности рассматриваемых временных рядов.

Обычно прогнозирование осуществляется с помощью соответствующих алгоритмов по нескольким последним реализациям ряда, образующим так называемое скользящее окно [6]. При этом если в окно попадают значения ряда, выходящие за пределы обычного интервала их изменения (существенный рост энергопотребления при непредвиденном увеличении числа потребителей энергии, значительное падение урожайности при сильной засухе, резкий рост заболеваемости населения при какой-либо эпидемии и пр.), то эти значения, называемые далее *особыми*, могут привести к большой погрешности в прогнозировании. Разумеется, влияние особых значений ряда на точность прогнозирования в определенной степени можно уменьшить за счет некоторого увеличения ширины окна, т.е. повышения степени сглаживания ряда. Однако при этом возрастает динамическая составляющая погрешности прогнозирования вследствие большей инерционности прогнозирующих алгоритмов.

В известной литературе вопросы прогнозирования временных рядов с указанными особыми значениями не анализировались. Поэтому авторы ставят задачу проиллюстрировать влияние особых значений ряда на результаты прогнозирования с помощью наиболее распространенных классических алгоритмов, предлагают метод модернизации этих алгоритмов для случая наличия в скользящем окне особых значений и показывают, что этот метод позволяет существенно улучшить точность прогнозирования.

Анализируемые алгоритмы и предлагаемый метод их модернизации

В работе исследовались три вида классических алгоритмов прогнозирования: скользящего среднего, линейные и параболические, представляемые согласно основному (базовому) методу или методу с использованием весовых коэффициентов. Анализ проводился для случая одношагового прогнозирования. Если, например, в скользящее окно попадают три последовательные значения X_i ряда, и линейный алгоритм строится с помощью метода наименьших квадратов, а прогнозирование производится на шаг t , где $t \geq 0$, то для основного метода значение прогноза Π_t находится следующим образом:

$$\Pi_t = \begin{cases} (X_{t-1} + X_{t-2} + X_{t-3})/3 & \text{— для алгоритма скользящего среднего,} \\ (4X_{t-1} + X_{t-2} - 2X_{t-3})/3 + (X_{t-1} - X_{t-3})t/2 & \text{— для линейного алгоритма,} \\ 3X_{t-1} - 3X_{t-2} + X_{t-3} + (5X_{t-1} - 8X_{t-2} + \\ + 3X_{t-3})t/2 + (X_{t-1} - 2X_{t-2} + X_{t-3})t^2/2 & \text{— для параболического алгоритма.} \end{cases}$$

Если применяется метод с использованием весовых коэффициентов, то для алгоритмов прогнозирования получаем следующие выражения:

$$P_t = \begin{cases} (k_1 X_{t-1} + k_2 X_{t-2} + k_3 X_{t-3}) / (k_1 + k_2 + k_3) & \text{— для алгоритма скользящего среднего,} \\ [(k_1 X_{t-1} + k_2 X_{t-2} + k_3 X_{t-3})(k_1 + 4k_2 + 9k_3) - \\ - (k_1 X_{t-1} + 2k_2 X_{t-2} + 3k_3 X_{t-3})(k_1 + 2k_2 + 3k_3) - \\ - (k_1 X_{t-1} + 2k_2 X_{t-2} + 3k_3 X_{t-3})(k_1 + k_2 + k_3)t + \\ + (k_1 X_{t-1} + k_2 X_{t-2} + k_3 X_{t-3})(k_1 + 2k_2 + 3k_3)t] / \\ / [(k_1 + k_2 + k_3)(k_1 + 4k_2 + 9k_3) - (k_1 + 2k_2 + 3k_3)^2] & \text{— для линейного алгоритма,} \\ 3k_1 X_{t-1} - 3k_2 X_{t-2} + k_3 X_{t-3} + 0,5(5k_1 X_{t-1} - 8k_2 X_{t-2} + \\ + 3k_3 X_{t-3})t + 0,5(k_1 X_{t-1} - 2k_2 X_{t-2} + k_3 X_{t-3})t^2 & \text{— для параболического алгоритма.} \end{cases}$$

Ввод в эти выражения весовых коэффициентов k_i , используемых в соответствующих случаях в методе наименьших квадратов [14], преследует цель уменьшить ошибки прогнозирования путем оптимизации значений k_i . Далее будет показано, что определенного результата таким путем действительно можно достичь.

Нетрудно увидеть, что выбирая $k_1 = k_2 = k_3 = 1$, получаем выражения для прогнозирования согласно основному методу, который, следовательно, является частным случаем метода с использованием весовых коэффициентов.

С целью дальнейшего улучшения точности прогнозирования авторы предлагают простой метод, который заключается в следующем: если в скользящее окно попадает особое значение ряда X_t , то в выражении для определения значения прогноза на основе данного окна X_t следует заменить на P_t . Следовательно, предлагаемый метод может дополнять большинство известных методов прогнозирования временных рядов и в сочетании с основным методом прогнозирования обеспечивает простоту реализации. Применение же метода весовых коэффициентов при расширении скользящего окна усложняет программы (с изменением местоположения особого значения в окне алгоритм прогнозирования изменяется).

Сравнительный анализ методов прогнозирования

Рассмотрим случай одношагового прогнозирования по трем значениям X_t в скользящем окне. Отметим, что для других, наиболее часто используемых окон, включающих в себя 2, 4 или 5 значений ряда, можно прийти к близким выводам.

В большинстве рассмотренных примеров особое значение X_t , под которым будем понимать значение, отличающееся на 30 и более процентов от прогноза, имеет место при $t = 2$. Для упрощения расчетных выражений примем, что прогноз осуществляется на шаг $t = 0$. Тогда для метода весовых коэффициентов в выражениях для линейного и параболического алгоритмов исчезнут два последних слагаемых.

Проведем качественное сравнение результатов прогнозирования в рассматриваемой задаче с помощью приведенных методов и алгоритмов. Для этого на рис.1-5 приводятся соответствующие диаграммы значений X_t и P_t . Во всех диаграммах на каждом шаге t принят следующий порядок следования величин (слева направо):  — значения X_t ,  — значения P_t , полученные согласно основному методу,  — значения P_t , полученные с использованием весовых коэффициентов,  — значения P_t , полученные с помощью модернизированного основного метода. В методе с использованием весовых коэффициентов значения k_i для особых значений приняты равными 0, а для остальных — равными 1.

Проиллюстрируем вначале на диаграммах влияние на P_t особого значения X_2 для двух характерных участков временного ряда: с постоянными и с линейно-изменяющимся X_t (рис.1-4).

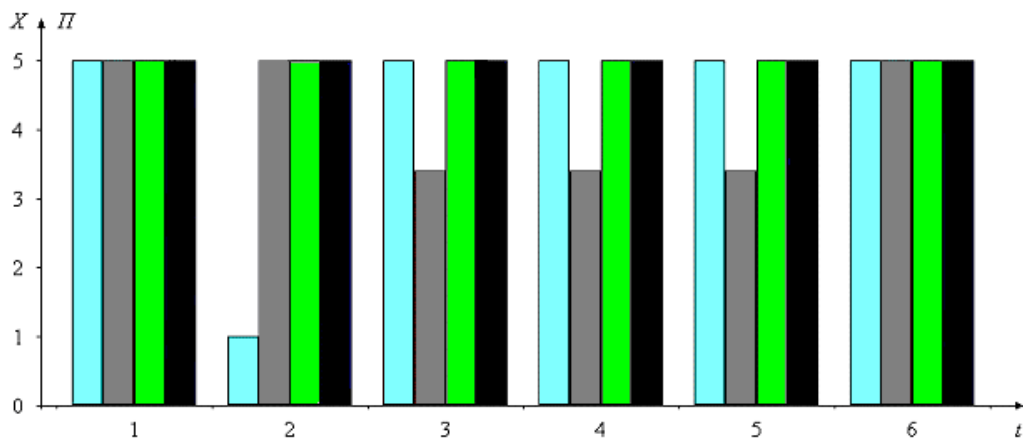


Рис.1. Диаграмма значений P_t , получаемых на основе скользящего среднего в случае $X_t = 5$ при $t \neq 2$ и $X_2 = 1$

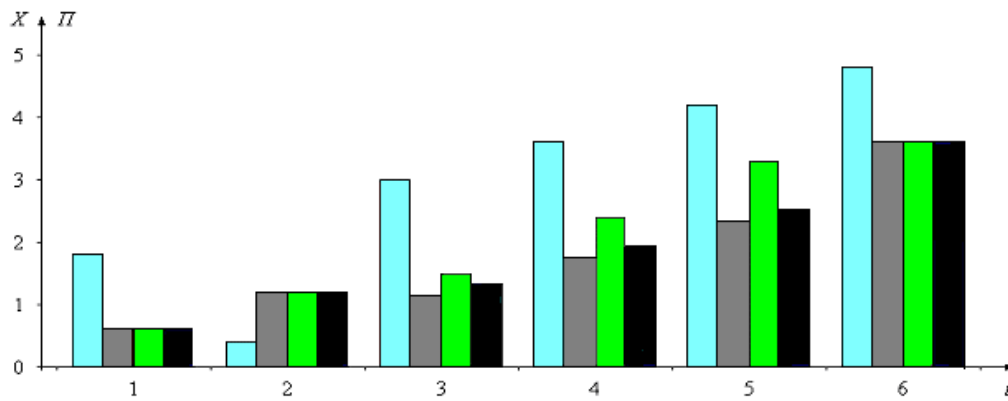


Рис.2. Диаграмма значений P_t , получаемых на основе скользящего среднего в случае $X_t = 1,2 + 0,6t$ при $t \neq 2$ и $X_2 = 0,4$

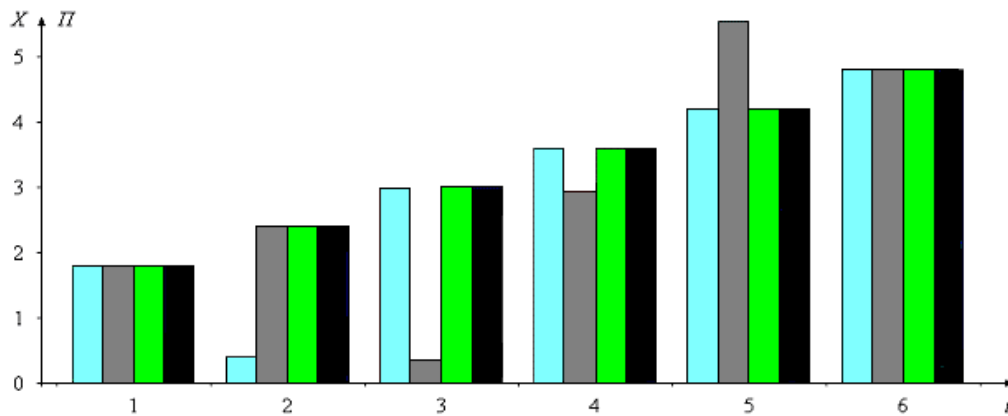


Рис.3. Диаграмма значений P_t , получаемых с помощью линейного алгоритма в случае $X_t = 1,2 + 0,6t$ при $t \neq 2$ и $X_2 = 0,4$

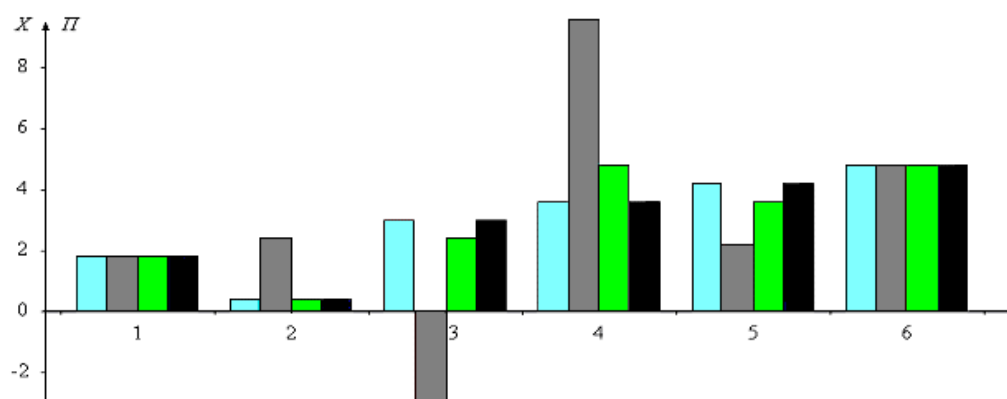


Рис.4. Диаграмма значений Π_t , получаемых с помощью параболического алгоритма в случае $X_t = 1,2 + 0,6t$ при $t \neq 2$ и $X_2 = 0,4$

Приведенные примеры облегчают понимание структуры последовательностей Π_t при $t \in [3, 5]$ и позволяют заключить, что при прогнозировании временных рядов с особыми значениями могут иметь место весьма грубые ошибки.

На рис.5 представлены диаграммы фрагмента ряда значений урожайности зерновых культур СССР в центнерах с гектара в первоначально оприходованном весе [6] и значений Π_t , полученных с помощью линейного алгоритма. Значение урожайности X_{1975} (10,9 ц/га) отличается от Π_{1975} (17,3 ц/га) в 1,6 раза. Поэтому при вычислении значений Π_{1976} , Π_{1977} и Π_{1978} согласно предложенному методу вместо X_{1975} использовалось Π_{1975} . Максимальные и средние значения погрешностей прогнозирования на отрезке 1976 — 1978-й гг. соответственно составили (в ц/га): 9,6 и 4,8 для основного метода, 6,5 и 4,5 для метода с использованием весовых коэффициентов, 4,4 и 3 для предложенного метода.

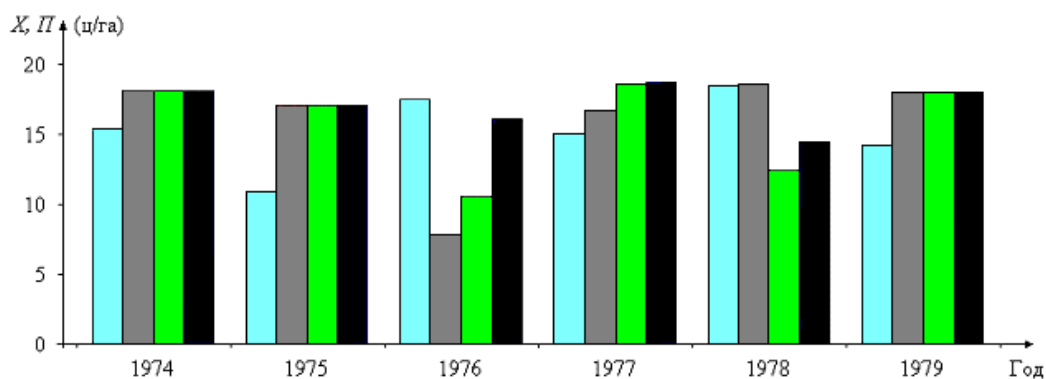


Рис. 5. Диаграмма значений урожайности зерновых культур в СССР и ее прогнозов

Приведенные диаграммы поясняют результаты прогнозирования с помощью разных методов и алгоритмов в условиях наличия особых значений в скользящем окне. Для получения необходимых количественных оценок точности прогнозирования было проведено моделирование 100000 участков ряда с семью значениями, среднее из которых являлось особым, а три последние значения сравнивались со значениями прогнозов, рассчитанных на основе указанного скользящего окна. Поскольку в окно попадали три значения ряда, то общее число указанных сравнений составило 300000. Моделируемые участки ряда принадлежали случайной последовательности $\{X_t\}$ с математическим ожиданием $M(X_t) = 0,5 + 0,1\sin(0,01t)$ и

одномерной функцией плотности типа «нормального колокола» на отрезке $[M(X_t) - 0,1, M(X_t) + 0,1]$ со средним квадратическим отклонением 0,0067. Особые значения являлись значениями случайной величины, равномерно распределенной в $[-1, -0,9]$ и в $[0,9, 1]$, т.е. эти значения всегда отстояли от неособых значений не менее чем на 0,2.

Приведенные в таблице результаты указанного моделирования дают представление о примерном соотношении точностей прогнозирования с помощью основного (осн.), с весовыми коэффициентами (вес.) и модернизированного основного (мод.) методов и трех анализируемых алгоритмов. Для рассмотренного ряда при изменении периода и амплитуды синусоиды в пределах 50% соотношение получаемых погрешностей почти не изменялось. При этом в случае использования алгоритмов скользящего среднего и линейного предложенная модернизация всегда существенно уменьшала максимальную и среднюю погрешности основного метода прогнозирования и обычно обеспечивала лучшую точность, чем метод с использованием весовых коэффициентов.

Значения абсолютных погрешностей прогнозирования

Алгоритм Метод	Скользящего среднего			Линейный			Параболический		
	Осн.	Вес.	Мод.	Осн.	Вес.	Мод.	Осн.	Вес.	Мод.
Максимальная погрешность	0,610	0,161	0,156	2,089	0,472	0,293	4,737	1,132	1,612
Средняя погрешность	0,241	0,031	0,031	0,548	0,062	0,048	1,641	0,150	0,226

Выводы

Предложенный метод прогнозирования временных рядов отличается простотой и, дополняя основной метод прогнозирования, позволяет значительно уменьшать погрешности прогнозирования.

1. Гноенский Л.С., Каменский Г.А., Эльсгольц Л.Э. Математические основы теории управляемых систем. М.: Наука, 1989. 512 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. Вып.1. 288 с.; Вып.2. 197 с.
3. Семенов Н.А. Программы регрессионного анализа и прогнозирования временных рядов. Пакеты ПАРИС и МАВР. М.: Финансы и статистика, 1990. 111 с.
4. Ширяев А.Н. Основы стохастической финансовой математики. Т.1. Факты. Модели; Т.2. Теория. М.: ФАЗис, 1998. 1017 с.
5. Granger C.W.J., Newbold P. Forecasting Economic Time Series, 2nd ed. Academic Press, Inc., 1986. 388 p.
6. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э.Фигурнова. М.: ИНФРА-М, 1998. 528 с.
7. Бачманов А.А. // Вестник НовГУ. Сер.: Математика и информатика. 2002. № 22. С. 46-49.
8. Кирьянов Б.Ф., Бачманов А.А. // Материалы научной сессии ННЦ СЗО РАМН. Т.2. М.: Медицина, 2003. С.76-82.
9. Бриллинджер Д. Временные ряды. М.: Мир, 1980. 536 с.
10. Кендэлл М. Временные ряды. М.: Финансы и статистика, 1976. 440 с.
11. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. М.: Статистика, 1979. 254 с.
12. Главные компоненты временных рядов: метод «Гусеница» / Под ред. Д.Л.Данилова и А.А.Жиглявского. СПб: СПбГУ, 1997. 308 с.
13. Golyandina N., Nekrutkin V., Solntsev V. Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall/CRC, 2001. 182 p.
14. Бахвалов Н.С. Численные методы. Т.1. М.: Наука, 1975. 632 с.