УДК 519.6
ББК 22.19
Б 94

Бучацкая В.В.
Кандидат технических наук, доцент кафедры прикладной математики и информационных технологий факультета математики и компьютерных наук Адыгейского государственного университета, тел. (8772) 59-37-32, e-mail: buch_vic@mail.ru

Методика определения интервальных оценок при прогнозировании методами экстраполяции

(Рецензирована)

Аннотация
Определение доверительных интервалов прогноза является одной из основных задач, возникающих при экстраполяции. На основе сведений из литературных источников установлена связь между величиной доверительного интервала и длиной временного ряда, а также периодом упреждения прогноза. Проанализированы результаты расчета величины доверительного интервала для заданного временного ряда.
Ключевые слова*: метод прогнозирования, экстраполяция, тренд, период упреждения, средняя квадратическая ошибка.*

Buchatskaya V.V.
Candidate of Technical Sciences, Associate Professor of the Department of Applied Mathematics and Information Technology, Faculty of Mathematics and Computer Science, Adyghe State University, ph. (8772) 59-37-32, e-mail: buch_vic@mail.ru

Technique to define the interval estimates when forecasting by extrapolation methods

Abstract
Definition of confidential intervals of a forecast is one of the main objectives arising at extrapolation. On the basis of data from literature sources a connection is established between the size of a confidential interval and the length of a temporary row, as well as the period of anticipation of a forecast. Results of calculation of the size of a confidential interval for the set temporary row are analyzed.
Keywords: *forecasting method, extrapolation, trend, anticipation period, average quadratic error.*

Базовые алгоритмы экстраполяция дают точечную прогностическую оценку. Однако, точное совпадение фактических данных и прогностических точечных оценок, полученных с их помощью, является маловероятным. Следовательно, такой оценки недостаточно и необходимо получение интервальной оценки. В этом случае прогноз, охватывая некоторый интервал значений прогнозируемой переменной, является более надежным. Определение доверительных интервалов прогноза является одной из основных задач, возникающих при экстраполяции. Это связано с тем, что расхождение между прогнозным и фактическим значениями имеет следующие источники [1]:

1) выбор формы кривой, характеризующий тренд, содержит элемент субъективизма, основан на опыте работы системного аналитика с методами прогнозирования, при этом с уверенностью нельзя утверждать, что выбранная форма кривой является единственно возможной или наилучшей для экстраполяции в данных конкретных условиях;

2) оценивание параметров кривых (тренда) производится на основе ограниченной совокупности наблюдений, каждое из которых содержит случайную компоненту; в силу этого параметрам кривой, а следовательно, и ее положению в пространстве свойственна некоторая неопределенность;

3) тренд характеризует некоторый средний уровень ряда на каждый момент времени; отдельные наблюдения, как правило, отклонялись от него в прошлом; логично ожидать, что подобного рода отклонения будут происходить и в будущем.

Погрешность, связанная со вторым и третьим источником, может быть отражена в виде доверительного интервала прогноза при принятии некоторых допущений о свойстве ряда. С помощью такого интервала точечный экстраполяционный прогноз преобразуется в интервальный.

В основу расчета доверительного интервала прогноза может быть положен измеритель колеблемости ряда наблюдаемых значений признака. Чем выше эта колеблемость, тем менее определено положение тренда в пространстве «уровень – время» и тем шире должен быть интервал для вариантов прогноза при одной и той же степени доверия. Следовательно, вопрос о доверительном интервале прогноза следует начать с рассмотрения измерителя колеблемости. Обычно такой измеритель определяют в виде среднего квадратического отклонения (стандартного отклонения) фактических наблюдений от расчетных, полученных при выравнивании динамического ряда.

В общем виде доверительный интервал для тренда определяется по формуле ([1, с. 162]):

$$\hat{y}_t \pm t_\alpha s_{\hat{y}}, \tag{1}$$

где $s_{\hat{y}}$ – средняя квадратическая ошибка тренда;

$\hat{y}_t$ – расчетное значение y_t ;

t_α – значение t – статистики Стьюдента.

Если $t=t+L$, то формула (1) определит значение доверительного интервала для тренда, продленного на L единиц времени.

Известно, что доверительный интервал определяется по формуле ([1, с. 164]):

$$\hat{y}_{t+L} \pm s_y K^*, \qquad K = \sqrt{\frac{n+1}{n} + \frac{3(n+2L-1)^2}{n(n^2-1)}},$$

где $K^*=t_\alpha K$.

Значение K зависит только от продолжительности наблюдения и периода упреждения. При увеличении продолжительности наблюдения (n) значения K уменьшаются, наоборот, с ростом величины L они растут.

В практике встречаются случаи, когда для экстраполяции можно применить несколько типов кривых [2]. В [1, с. 172, 174] получены выражения для доверительных интервалов прогноза полиномов 1, 2 и 3 степени, а также для уравнения экспоненты и логарифмической параболы (табл. 1).

Как видно из приведенных формул при одной и той же величине s_y доверительный интервал прогноза тем шире, чем выше степень полинома, характеризующего тренд. Это объясняется тем, что дисперсия уравнения тренда определяется как взвешенная сумма дисперсий соответствующих параметров уравнений. Если тренд лучше описывается кривой более высокого порядка, то соответственно квадратическая ошибка $s_{\hat{y}}$ будет ниже и, следовательно, доверительный интервал уже, чем, например, при линейном тренде.

На рисунке 1 приведены диаграммы, характеризующие зависимость величины K от n при условии, что L=1, L=5, L=7, L=10. При увеличении продолжительности наблюдения (n) значения K и K^* уменьшаются, наоборот, с ростом величины L они растут.

Зависимость K от L для n=5, n=10, n=15 и n=20 показана на рисунке 2. Как видно из приведенных графиков, влияние периода упреждения неодинаково для различных значений n: чем больше длительность наблюдения, тем меньшее влияние она оказывает на период упреждения. Характер влияния изменится в том случае, когда L определяется как относительная величина от n.

Отметим, что n влияет на величину доверительного интервала не только через K. Значение статистики t_α при фиксированном значении вероятности определяется, как известно, числом степеней свободы. Последнее же зависит от n.

Таблица 1

Расчетные формы доверительных интервалов различных моделей

Тип модели	Доверительный интервал
Линейная	$s_y K^* = t_\alpha s_y \sqrt{1 + \frac{1}{n} + \frac{1}{\sum t^2} t_L^2}$
Полином 2-го порядка	$s_y K^* = t_\alpha s_y \sqrt{1 + \frac{1}{\sum t^2} t_L^2 + \frac{\sum t^4 - \left(2\sum t^2\right) t_L^2 + n t_L^4}{n\sum t^4 - \left(\sum t^2\right)^2}}$
Полином 3-го порядка	$s_y K^* = t_\alpha s_y \sqrt{1 + \frac{1}{\sum t^2} t_L^2 + \frac{\sum t^4 - \left(2\sum t^2\right) t_L^2 + n t_L^4}{n\sum t^4 - \left(\sum t^2\right)^2} + \frac{\left(\sum t^6 - 2\sum t^4\right) t_L^2 + \left(\sum t^2\right) t_L^6}{\sum t^2 \sum t^6 - \left(\sum t^4\right)^2}}$
Экспонента	$ant\log(\log \hat{y}_{t+L} \pm s_y K^*)$, где $s_y K^* = t_\alpha s_y \sqrt{1 + \frac{1}{n} + \frac{1}{\sum t^2} t_L^2}$
Логарифмическая парабола	$ant\log(\log \hat{y}_{t+L} \pm s_y K^*)$, где $s_y K^* = t_\alpha s_y \sqrt{1 + \frac{1}{\sum t^2} t_L^2 + \frac{\sum t^4 - \left(2\sum t^2\right) t_L^2 + n t_L^4}{n\sum t^4 - \left(\sum t^2\right)^2}}$

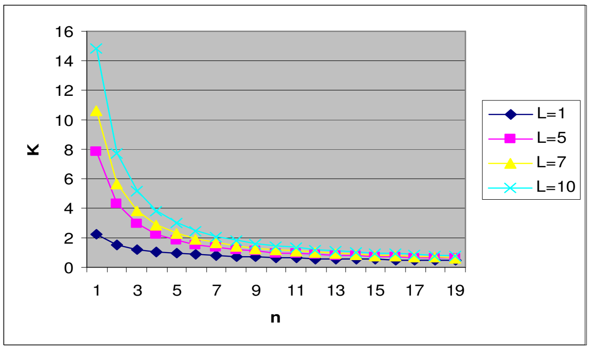

Рис. 1. Зависимость K от числа наблюдений n

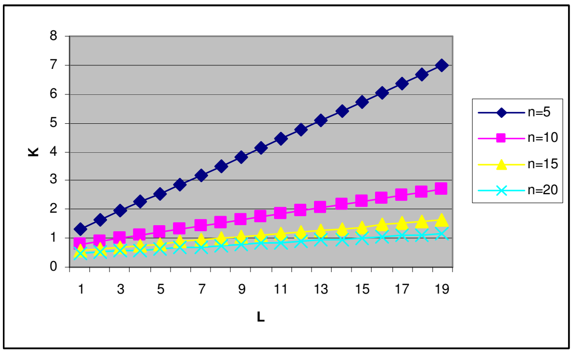

Рис. 2. Зависимость K от продолжительности периода упреждения L

Определим доверительный интервал для прогнозирования показателя, используемого при оценке эффективности деятельности органов местного самоуправления городских округов и муниципальных районов (Краснодарский край). Данные использованы из ежегодного статистического сборника 2010–2011 гг. [3].

Исследуем результаты прогнозирования значений ряда, состоящего из 12 значений (n=12). Построим следующие трендовые модели прогнозирования: линейную, полином 2-й степени, полином 3-й степени (табл. 2) [4].

Как видно из таблицы 3, для полинома 2-й степени доверительный интервал оказался почти неизменным. Минимальная величина доверительного интервала получена для линейного тренда. Для полиномов 1-й и 3-й степени с ростом периода упреждения отношение ширины доверительного интервала к расчетному уровню возрастает, хотя и незначительно. Это объясняется тем, что ряд возрастающий и по мере роста *t* числитель и знаменатель увеличиваются равномерно. Кроме того, в исследуемом ряду отсутствуют аномальные выбросы. В том случае, когда в результате расчетов доверительный интервал оказывается достаточно широким, его можно сократить, уменьшив доверительную вероятность получаемых оценок.

Таблица 2

Модели прогнозирования

№ п/п	Обозначение	Описание модели		Параметры модели
1.	LR(1)	Линейная	$U_t = a_0 + a_1 t$	a_0=17,435; a_1=0,7453
2.	LR(2)	Полином 2-й степени	$U_t = a_0 + a_1 t + a_2 t^2$	a_0=17,452; a_1=0,738; a_2=0,0006
3.	LR(3)	Полином 3-й степени	$U_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$	a_0=17,452; a_1=0,738; a_2=0,0006

По изложенным ранее формулам из таблицы 1 рассчитаем доверительные интервалы этих моделей (табл. 3).

Таблица 3

Результаты расчета доверительных интервалов (доверительная вероятность 90%)

№ п/п	Модель	*L*=1		*L*=2		*L*=3		*L*=4	
		$\hat{y}_t$	Доверительный интервал	$\hat{y}_t$	Доверительный интервал	$\hat{y}_t$	Доверительный интервал	$\hat{y}_t$	Доверительный интервал
1.	LR(1)	26,485	$\hat{y}_t \pm 1{,}15$	27,24	$\hat{y}_t \pm 1{,}3$	27,994	$\hat{y}_t \pm 1{,}46$	28,748	$\hat{y}_t \pm 1{,}62$
2.	LR(2)	26,399	$\hat{y}_t \pm 3{,}33$	27,174	$\hat{y}_t \pm 3{,}32$	27,901	$\hat{y}_t \pm 3{,}32$	28,657	$\hat{y}_t \pm 3{,}32$
3.	LR(3)	25,948	$\hat{y}_t \pm 4{,}02$	25,895	$\hat{y}_t \pm 4{,}13$	25,468	$\hat{y}_t \pm 4{,}41$	24,617	$\hat{y}_t \pm 4{,}03$

Примечания:

1. Четыркин Е.М. Статистические методы прогнозирования. 2-е изд., перераб. и доп. М.: Статистика, 1977. 199 с.
2. Симанков В.С., Бучацкая В.В. Выбор методов прогнозирования при исследовании сложных систем // Вестник Адыгейского государственного университета. Сер. Естественно-математические и технические науки. 2012. Вып. 2 (101). С. 114-119. URL: http://vestnik.adygnet.ru

References:

1. Chetyrkin E.M. Statistical methods of forecasting. 2nd ed., revised and enlarged. M.: Statistics, 1977. 199 pp.
2. Simankov V.S., Buchatskaya V.V. Choice of methods of forecasting in researches of complicated systems // The Bulletin of the Adyghe State University. Series Natural-Mathematical and Technical Sciences. 2012. Iss. 2 (101). P. 114-119. URL: http://vestnik.adygnet.ru

3. Территориальный орган федеральной службы государственной статистики по Краснодарскому краю. URL: http://www.krsdstat.ru/munstat/default.aspx

4. Бучацкая В.В. Сравнительный анализ некоторых моделей прогнозирования временных рядов // Актуальные проблемы прикладной математики, информатики и механики: материалы междунар. конф. Воронеж, 2012. С. 8-10.

3. Territorial body of the Federal Service of State statistics of Krasnodar region. URL: http://www.krsdstat.ru/munstat/default.aspx

4. Buchatskaya V.V. Comparative analysis of some models of forecasting of time series // Actual problems of applied mathematics, informatics and mechanics: materials of international conference. Voronezh, 2012. P. 8-10.