УДК 338.27+519.86

МОДИФИЦИРОВАННЫЙ МЕТОД ЭКСТРАПОЛЯЦИОННОГО ПРОГНОЗИРОВАНИЯ

Г. А. НАРЫШЕВ, Н. В. ПАРХОМЕНКО
Гомельский государственный технический университет имени П.О. Сухого, Республика Беларусь

Методы экстраполяции тенденций являются, пожалуй, самыми распространенными и наиболее разработанными среди всей совокупности методов прогнозирования. Экстраполяционные методы прогнозирования основной упор делают на выделение наилучшего в некотором смысле описания тренда и на определении прогнозных значений путем его экстраполяции[2]. Модификацией данного метода может служить экстраполяционное прогнозирование на основе определения оптимального числа моментов предыстории.

Суть предлагаемого метода сводится к тому, что среди всех рассматриваемых кривых для целей прогнозирования отбирается та, которая дает лучшие результаты при прогнозировании, а не при описании предыстории.

Эта идея реализуется последовательной проверкой прогнозирующей способности всех кривых, принятых для исследования, причем в процессе проверки находится такое оптимальное число моментов предыстории, на основе учета которых получается наилучший прогноз для каждой кривой.

Рассмотрим алгоритм метода более подробно.

Пусть исходная информация задана динамическим рядом показателя X за t временных интервалов:

$$x_1, x_2, x_3, \dots, x_t \qquad (1)$$

и требуется рассчитать прогнозное значение этого показателя в период $t+l$, то есть, x_{t+l}. Кроме того, необходимо определить возможную ошибку прогноза.

Решение задачи можно разложить на ряд процедур.

Процедура 1. Из исходного динамического ряда формируется $t-l-3$ серий вспомогательных рядов последовательно из $t-1, t-2, t-3...$ членов с соблюдением порядка членов в исходном динамическом ряду следующим образом:

исходный ряд $x_1 x_2 \dots x_{t-l} \dots x_t$

I серия рядов $x_1 x_2 \dots x_{t-l-1} \dots x_{t-1}$

(2 ряда) $x_2 x_3 \dots x_{t-l} \dots x_t$

II серия рядов $x_1 x_2 \dots x_{t-l-2} \dots x_{t-2}$

(3 ряда) $x_2 x_3 \dots x_{t-l-1} \dots x_{t-1}$

$x_3 x_4 \dots x_{t-l} \dots x_t$

………………..

k–я серия рядов $x_1 x_2 \dots x_{t-l-k} \dots x_{t-k}$

((k+1) рядов) $x_2x_3\ldots x_{t-l-k+1}\ldots x_{t-k+1}$

……………………

i-тый ряд $x_ix_{i+1}\ldots x_\tau\ldots x_{\tau+l}$

…………………….

$x_{k+1}x_{k+2}\ldots x_{t-l}\ldots x_t$,

где $i=1,2,...,k+1$; $k=1,2,...,t-l-3$; $\tau=t-l-k+(i-1)$

……………………

$(t-l-2)$ серия рядов $x_1x_2\ldots x_{2+l}$

$((t-l-1)$ ряда) $x_2x_3\ldots x_{3+l}$

……………………

$x_{t-l-2}x_{t-l-1}\ldots x_t$.

При этом $(t-l-2)$ серия рядов формируется только для кривых 1-го порядка. Для кривых 2-го порядка формируется $(t-l-3)$ серии рядов.

Процедура 2. Для каждого i-го ряда k-той серии ($i=1,2,...,k+1$; $k=1,2,...,t-l-3$) по первым τ членам ($\tau=t-l-k+(i-1)$) этого ряда методом наименьших квадратов находятся параметры a и b уравнений прямых $X=a+b\cdot t$.

Процедура 3. Используя полученные уравнения, рассчитываем по каждому i-му ряду k-той серии прогнозные значения X на $(\tau+l)$ период. Обозначим эти прогнозные значения через $\overline{X}_{ik}^{\tau+l}$.

Процедура 4. Вычисляются ошибки экстраполяции по уравнениям прямых для всех рядов по следующей формуле:

$$\delta_{ik}=\left|\frac{x_{ik}^{\tau+l}-\overline{x}_{ik}^{\tau+l}}{x_{ik}^{\tau+l}}\right|, \tag{2}$$

где $x_{ik}^{\tau+l}$ - фактическое значение показателя X, стоящее последним в i-м ряду k-той серии.

Процедура 5. Вычисляются средние ошибки экстраполяции для каждой серии рядов по следующей формуле:

$$\delta_k=\frac{1}{k+1}\sum_{i=1}^{k+1}\delta_{ik}\ . \tag{3}$$

Процедура 6. Отыскивается $\min_k \delta_k=\delta_{k'}$. Номер серии k' укажет оптимальное число моментов предыстории, равное $t-l-k'$, при котором ошибка экстраполяции минимальна.

Процедуры 2-6 повторяются последовательно для всех других исследуемых кривых. Предположим, что число таких кривых p. В результате повторений процедур 2-6 для каждой из этих кривых получим p значений ошибок экстраполяции $\delta_{k'}^{p}$.

Процедура 7. Отыскивается $\min_p \delta_{k'}^{p}=\delta_{k'}^{p'}$. Номер кривой p' укажет нам наилучшую прогнозирующую кривую.

Процедура 8. Находится доверительный интервал Δ_δ для $\delta_{k'}^{p'}$ по следующей формуле:

$$\Delta_\delta = \pm \frac{S}{\sqrt{k'+1}} \cdot T\,, \qquad (4)$$

где S – среднее квадратическое отклонение ошибок $\delta_{ik'}^{p'}$ от $\delta_{k'}^{p'}$;

T – статистика Стьюдента.

Процедура 9. По $t-l-k'$ последним членам исходного ряда методом наименьших квадратов находятся параметры наилучшей прогнозирующей кривой p'.

Процедура 10. Полученное в результате выполнения процедуры 9 уравнение используется для расчета прогнозного значения x_{t+l}.

Процедура 11. Рассчитывается доверительный интервал прогнозного значения x_{t+l} по следующей формуле:

$$\Delta_x = \pm\left(\delta_{k'}^{p'} + \Delta_\delta\right) \cdot x_{t+l}\,. \qquad (5)$$

Смысл описанных процедур заключается в опробовании и оценке точности прогнозирования по двум, трем и более моментам предыстории последовательно на данных всего динамического ряда. Средние ошибки прогноза по одному и тому же числу моментов предыстории (по каждой серии рядов) служат критерием точности прогноза, а наименьшая средняя ошибка соответствует оптимальному числу моментов, на основе учета которых прогноз получается наилучшим. В то же время сравнение средних ошибок прогнозирования, соответствующих оптимальному числу моментов предыстории по различным кривым, позволяет осуществить выбор наилучшего прогнозирующего уравнения.

За доверительный интервал принимается сумма средней ошибки экстраполяции, рассчитанной для соответствующего (оптимального) числа моментов предыстории, и ошибки оценки этой средней. При этом делается допущение о том, что значение средней ошибки в будущем не изменится. В связи с трудностями теоретического обоснования вероятности непопадания фактических значений в доверительный интервал прогноза может быть предложен экспериментальный метод установления частоты такого непопадания. Для этого необходимо провести достаточно большое число испытаний метода на материалах различных динамических рядов и сравнить полученные прогнозы (с учетом доверительных интервалов) с известными фактическими значениями. Отношение числа фактических значений, не попавших в доверительный интервал, к общему числу испытаний и даст нам оценку правильности выбора величины доверительного интервала. Этим способом можно рассчитать также и величину доверительного интервала, соответствующего, например, 95 или 99% вероятности попадания в него фактического значения исследуемого показателя.

Недостатком предложенного метода является невозможность получения достоверных статистических выводов о прогнозирующих способностях кривых в тех случаях, когда предыстория задана короткими динамическими рядами, или в случаях, когда число членов динамического ряда в серии близко к числу членов исходного ряда. В этих случаях выводы о прогнозирующих способностях кривых основаны на малом числе испытаний, поэтому доверительный интервал прогноза увеличивается на величину ошибки оценки средней ошибки экстраполяции, что в некоторой степени позволяет учесть это обстоятельство. Полностью устранить этот недостаток можно только увеличением протяженности исходного динамического ряда.

При использовании данного метода действительны все требования, которые предъявляются к исходным данным (сопоставимость во времени и др.) при использовании других методов экстраполяционного прогнозирования.

Предложенный метод является развитием идеи прогноза методом скользящей средней с поиском оптимальной длины ряда, на основе которого рассчитывается эта средняя [1].

Модифицированный метод экстраполяционного прогнозирования основан на испытании прогностических способностей различных кривых. Современные средства вычислительной техники позволяют быстро и качественно выполнить все эти расчеты, при этом мы не ограничены в каких-либо пределах, что позволяет испытывать большое число таких кривых. Практические расчеты показали, что, за редким исключением, наилучшими кривыми являются кривые, порядок которых не превышает двух. Этим как раз доказывается обоснованность сомнений исследователей в прогнозирующих способностях кривых высоких порядков. Многочисленные испытания показали, что наилучшим набором кривых является следующий:

$$X = a + b \cdot t \tag{6}$$

$$X = a + \frac{b}{t} \tag{7}$$

$$\frac{1}{X} = a + \frac{b}{t} \tag{8}$$

$$X = a + b \cdot \ln t \tag{9}$$

$$\ln X = a + b \cdot \ln t \tag{10}$$

$$\ln X = a + b \cdot t \tag{11}$$

$$X = a + b \cdot t + c \cdot t^2 \tag{12}$$

$$X = a + \frac{b}{t} + \frac{c}{t^2} \tag{13}$$

$$\frac{1}{X} = a + \frac{b}{t} + \frac{c}{t^2} \tag{14}$$

$$X = a + b \cdot \ln t + c \cdot \ln^2 t \tag{15}$$

В некоторых случаях этот набор целесообразно дополнить уравнением

$$x_{t+1} = \frac{1}{n}\sum_{i=0}^{n-1} x_{t-i} \tag{16}$$

для прогноза по скользящей средней, причем протяженность скользящей средней (n) берется каждый раз оптимальной (в соответствии с вышеизложенным алгоритмом).

При анализе многих динамических рядов удалось выявить три типа зависимости точности прогноза от длины предыстории:

1. Точность прогнозирования уменьшается с ростом длины предыстории – это наиболее распространенный тип, он характеризует процессы с часто меняющимися темпами развития объекта.

2. Точность прогнозирования растет с ростом длины предыстории. Процессы, характерные для такой зависимости носят стабильный, устойчивый характер на протяжении длительного периода времени, а встречающиеся в этом периоде отклонения от тренда носят случайный характер.

3. Точность прогнозирования сначала растет с ростом длины предыстории, а потом, по достижении некоторой максимальной величины, начинает падать. Процессы развития объекта, характеризующиеся этим типом зависимости, занимают, по-видимому, промежуточное положение между первыми двумя.

Выявленные зависимости показывают, что длина предыстории, учитываемая при прогнозе, очень сильно влияет на точность прогнозирования (иногда ошибки прогноза при учете разной предыстории отличаются на порядок), поэтому нахождение этой зависимости в каждом конкретном случае должно стать обязательным. Игнорирование этого обстоятельства было одним из источников больших ошибок при прогнозе, что вызывало настороженное отношение исследователей к экстраполяционным методам прогнозирования вообще.

Отдельные результаты испытаний метода противоречат некоторым сложившимся представлениям о предпосылках экстраполяционного прогнозирования. Так, всегда полагалось, что прогноз должен осуществляться по предыстории, по крайней мере, не меньшей, чем период упреждения, а часто выдвигалось требование превышения числа моментов предыстории над периодом упреждения в несколько раз. У нас же нередки случаи, когда при прогнозе на 5 лет оптимальное число моментов предыстории составляло всего два года, причем прогноз при этом получался наиболее точным. Эта необычность выводов кажется таковой только на первый взгляд. По-видимому, для процессов, характеризующихся частой сменой темпов развития, прогноз по короткой предыстории вполне допустим, так как в этом случае учет более отдаленных моментов только исказит результаты прогнозирования. Подобные выводы не означают, однако, что прогноз вообще можно осуществлять по короткой предыстории.

Чтобы обосновать такую возможность, необходимо иметь достаточно протяженный исходный ряд и чем протяженнее будет этот ряд, тем более достоверными будут выводы об оптимальном числе моментов предыстории.

Результаты сравнительного анализа различных методов прогнозирования (испытывались методы экстраполяции тенденций, экспоненциального сглаживания, гармонических весов, авторегрессия и модифицированный метод экстраполяции) показали, что предложенный метод имеет целый ряд преимуществ перед всеми другими, главное из которых – более высокая точность прогнозов.

Некоторые исследователи возражают против использования методов экстраполяционного прогнозирования (имеется в виду экстраполяция в узком смысле) для прогнозов на большие периоды упреждения (5 лет и более). Нам кажется, что при выборе метода необходимо подходить к этому вопросу диалектически: бывают случаи, когда экстраполяционные прогнозы дают хорошие результаты и на периоде в 20 лет, бывают и такие, когда экстраполяция всего на два года приводит к абсурдным результатам. Видимо, следует выявить признаки такого различия и затем применять аппарат экстраполяционного прогнозирования только в благоприятных случаях. В качестве одного из таких признаков, способного указывать на возможность достоверных прогнозов, является такая характеристика динамического ряда, как «гладкость» этого ряда. Наиболее приемлемым, на наш взгляд, показателем «гладкости» можно считать остаточную дисперсию отклонений фактических точек динамического ряда от соответствующих точек кривой первого или второго порядка, взятой из достаточно представительного набора таких кривых и наилучшим образом описывающей исследуемый процесс. Вместо остаточной дисперсии можно рекомендовать также среднее абсолютное отклонение фактических точек от выравненных по кривой, отобранной как раз по этому критерию из всех исследуемых. Необходимость привлечения большого первоначального набора кривых диктуется различиями в характере развития того или иного экономического процесса, а ограничение порядка кривых накладываются из тех соображений, что кривые высоких порядков могут хорошо аппроксимировать и случайные отклонения от нормального хода процесса.

Можно показать, что при прогнозе на основании динамических рядов определенной «гладкости» экстраполяционные прогнозы получаются с определенными ошибками. Следовательно, определив «гладкость» ряда, мы всегда можем оценить возможную ошибку экстраполяционного прогноза. Отсюда вытекает возможность установления границы применимости методов экстраполяционного прогнозирования в том случае, если нам заранее задана возможная погрешность прогнозов. Если выводы о величине возможной ошибки окажутся неблагоприятными, аппарат экстраполяционного прогнозирования применять нецелесообразно.

Наряду с «гладкостью» динамических рядов на точность прогнозирования большое влияние оказывает также длина учитываемой предыстории и период упреждения. Чем длиннее предыстория, тем более надежные выводы можно сделать о будущем, и, наоборот, с увеличением периода упреждения достоверность этих выводов падает. Учитывая эти соображения, можно провести исследование на предмет определения сфер применимости всех известных методов экстраполяционного прогнозирования, и это исследование даст нам возможность заранее, не производя прогностических расчетов, указать наиболее точный метод и оценить возможную ошибку прогнозирования, исходя только их таких известных величин, как «гладкость» ряда, его протяженность и необходимая величина периода упреждения. В процессе этого исследования необходимо будет также уточнить границы доверительных интервалов прогноза. Это можно сделать эмпирически, рассчитав уравнение вида

$$\Delta = f\left(\overline{\delta}\right), \tag{17}$$

где Δ - величина доверительного интервала прогноза;

$\overline{\delta}$ - средняя ошибка прогнозирования.

Данные для расчета параметров уравнения можно получить на основании большого числа прогностических испытаний на рядах определенной «гладкости». Для рядов с другой «гладкостью», уравнения, естественно, будут отличными. Следует учесть, что величина интервала может быть определена со стандартными уровнями доверия: 0,05; 0,01 или даже 0,001. Достоверность получаемых выводов будет зависеть от количества проводимых испытаний.

Модифицированный метод экстраполяционного прогнозирования позволяет очень просто корректировать прогнозы, если в них есть систематическая ошибка, а это может значительно повысить точность прогнозирования. Корректировку лучше всего производить по методу, предложенному г.Тейлом в [3]. Строя уравнение «прогноз - реализация» на основании данных испытаний точности прогноза по какой-либо кривой для рядов одной серии ($\widehat{x}^* = f\left(x^*\right)$, где $\widehat{x}^*$ и x^* - соответственно, скорректированный и нескорректированный прогнозы), мы можем в дальнейшем использовать его для корректировки получаемого прогноза. Особенно большую пользу приносит такая корректировка при прогнозе на большие периоды упреждения, значительно снижая среднюю ошибку прогнозирования.

В заключение отметим, что идея, на которой основан модифицированный метод экстраполяционного прогнозирования, может быть положена в основу всех других методов экстраполяции.

Литература

1. Розин Б.Б., Нарышев Г.А., Воронов Ю.П. Прогнозирование показателей работы сталеплавильных агрегатов. – В сб.: Моделирование и программирование планово-экономических задач. –Новосибирск: НГУ, 1965.
2. Теория прогнозирования и принятия решений / Под ред. Саркисяна С.А. – М.: Высшая школа, 1977.

3. Тейл Т. Прикладное экономическое прогнозирование. – М.: Прогресс, 1970.
4. Тейл Т. Экономические прогнозы и принятие решений. – М.: Статистика, 1971.
5. Шуман Х. Прогнозирование в отрасли промышленности. – М.: Экономика, 1971.
6. Эйрес Р. Научно-техническое прогнозирование и долгосрочное планирование. –М.: Мир, 1971.