Айвазян С.А. Основы эконометрики - М.:Юнити, 2001

Прогнозирование, основанное на использование моделей временных рядов

Ряд наблюдений $x(t_1), x(t_2), \dots, x(t_n)$ анализируемой случайной величины $\xi(t)$, произведенных в последовательные моменты времени $t_1, t_2, \dots, t_n$, называется временным рядом. Члены временного ряда не являются статистически независимыми и не являются одинаково распределенными. Это значит, что нельзя распространять свойства и правила статистич6еского анализа случайной выборки на временные ряды. С другой стороны, взаимозависимость членов временного ряда создает свою специфическую базу для построения прогнозных значений анализируемого показателя (т.е. для построения оценок $\hat{x}(N+k)$ для неизвестных значений $x(N+k)$) по наблюденным значениям $x(1), x(2), \dots, x(N)$.

На временной ряд влияют следующие факторы:

- долговременные, формирующие общую тенденцию в изменении анализируемого признака $x(t)$. Обычно эта тенденция описывают с помощью той или иной неслучайной функции $f_{тр}(t)$. Это функцию называют функцией тренда или тренд;

- сезонные, формирующие периодически повторяющуюся в определенное время года колебания анализируемого признака;

- циклические, формирующие изменения анализируемого признака, обусловленные действием долговременных циклов демографической, экономической, астрофизической и другой природы;

- случайные, на поддающиеся учету и регистрации.

При прогнозировании на основе статистического анализа временных рядов следует учитывать следующие условия:

- требуемый горизонт l прогнозирования, т.е. на сколько временных тактов (l) вперед мы собираемся строить наш прогноз; при $l \le 3$ прогноз обычно называется краткосрочным, при $3 < l \le 6$ – среднесрочным, а при $l > 6$ – долгосрочным (приведенная здесь классификация условна: она может зависить от конкретного смысла одного такта времени – подразумевается ли под ним день, неделя, квартал, год и т.д.);

- длина анализируемого временного ряда (условно говоря, при $N \le 50$ ряд считается коротким, а при $N > 50$ – длинным);

- наличие или присутствие в анализируемом ряду сезонной составляющей или каких-либо «нестандартностей» (скачкообразных изменений в поведении тренда, слишком большой величины дисперсии случайных остатков и т.п.) .

2.1.1 Метод экспоненциального сглаживания (Брауна)

Пусть анализируемый временной ряд представлен в виде $x(\tau)\ (\tau = 1,2,...,t)$

$$x(\tau) = a_0 + \varepsilon(\tau) \tag{1}$$

где a_0 – неизвестный параметр, не зависящий от времени, а $\varepsilon(\tau)$ - случайный остаток со средним значением, равным нулю, и конечной дисперсией. Экспоненциально взвешенная скользящая средняя (ЭВСС) ряда $x(\tau)$ в точке с параметром сглаживания (параметров адаптации) $\lambda\,(0 < \lambda < 1)$ определяется формулой

$$\bar{x}_\lambda(t) = \frac{1-\lambda}{1-\lambda^t} \sum_{j=0}^{t-1} \lambda^j x(t-j) \tag{2}$$

которая дает решение задачи на минимум экспоненциального взвешенного критерия метода наименьших квадратов, а именно:

$$\bar{x}_\lambda(t) = \arg\min_a \sum_{j=0}^{t-1} \lambda^j \left[x(t-j) - a\right]^2 \tag{3}$$

Коэффициент сглаживания λ можно интерпретировать также как коэффициент дисконтирования, характеризирующий меру обесценивания наблюдения за единицу времени.

Для длинных временных рядов (с «бесконечным прошлым») формула (2) сводится к виду

$$\bar{x}_\lambda(t) = (1-\lambda)\sum_{j=0}^{\infty} \lambda^j x(t-j) \tag{4}$$

В соответствии с простейшим вариантом метода экспоненциального сглаживания прогноз $\hat{x}(t;1)$ для неизвестного значения $x(t+1)$ по известной до момента времени t траектории ряда $x(\tau)$ строится по формуле

$$\hat{x}(t;1) = \bar{x}_\lambda(t) \tag{5}$$

Где значение $\bar{x}_\lambda(t)$ определено формулой (2) (для короткого временного ряда) или формулой (4) (для длинного временного ряда).

Формула (5) удобна, тем, что при появлении следующего $(t+1)$–го наблюдения $x(t+1)$ пересчет прогнозирующей функции $\hat{x}(t+1;1) = \bar{x}_\lambda(t+1)$ производится с помощью простого соотношения

$$\bar{x}_\lambda(t+1) = \lambda\bar{x}_\lambda(t) + (1-\lambda)x(t+1) \tag{6}$$

Для вывода этого соотношения выпишем (4) для моментов времени t и $t+1$:

$$\bar{x}_\lambda(t+1) = (1-\lambda)\left[x(t+1) + \lambda x(t) + \lambda^2 x(t-1) + \ldots\right]$$

$$\bar{x}_\lambda(t) = (1-\lambda)\left[x(t) + \lambda x(t-1) + \lambda^2 x(t-2) + \ldots\right] \qquad (7)$$

Вычтя из первого соотношения второе, умноженное на λ, получим $\bar{x}_\lambda(t+1) - \lambda \bar{x}_\lambda(t) = (1-\lambda)x(t+1)$, что и доказывает справедливость (6).

Метод экспоненциального сглаживания можно обобщить на случаи полиноминальной неслучайной составляющей анализируемого временного ряда:

$$x(t+\tau) = a_0 + a_1\tau + \ldots + a_k\tau^k + \varepsilon(\tau) \qquad (8)$$

где $k \geq 1$. В соотношении (8) начальная точка отчета времени сдвинута в текущий момент времени, что облегчает дальнейшие выкладки и вычисления. Соответственно, в схеме простейшего варианта метода прогноз $\hat{x}(t;1)$ значения $x(t+1)$ будет оправдываться соотношениями (8) при $\tau = 1$ и (5):

$$\hat{x}(t;1) = \hat{x}(t+1) = \hat{a}_0^{(k)}(t;\lambda) + \hat{a}_1^{(k)}(t;\lambda) + \ldots + \hat{a}_k^{(k)}(t;\lambda) \qquad (9)$$

где оценки $\hat{a}_0^{(k)}(t;\lambda)$, $\hat{a}_1^{(k)}(t;\lambda)$, …, $\hat{a}_k^{(k)}(t;\lambda)$ определяются следующим образом:

$$\sum_{j=1}^{\infty} \lambda^j \left[x(t-j) - a_0 - a_1 j - \ldots - a_k j^k\right]^2 \xrightarrow{} \min_{a_0, a_1, \ldots a_k} \qquad (10)$$

Для случая линейного тренда, т. е $Ex(t+\tau) = a_0 + a_1\tau$.

$$\hat{a}_0^{(1)}(t;\lambda) = \bar{x}_\lambda(t) + (1-\lambda)\sum_{j=0}^{\infty} \lambda^j x(t-j)$$

$$\hat{a}_1^{(1)}(t;\lambda) = (1-\lambda)\bar{x}_\lambda(t) + (1-\lambda)^2 \sum_{j=0}^{\infty} (j+1)\lambda^j x(t-j) \qquad (11)$$

Так что прогноз $\hat{x}(t;1)$ будущего значения $x(t+1)$ при линейном тренде будет вычисляться в соответствии с (9)

$$\hat{x}(t;1) = \hat{x}(t+1) = \hat{a}_0^{(1)}(t;\lambda) + \hat{a}_1^{(1)}(t;\lambda) \quad (12)$$

Где коэффициенты $\hat{a}_0^{(1)}(t;\lambda), \hat{a}_1^{(1)}(t;\lambda)$ определяются по формулам (10).

Пересчет коэффициентов $\hat{a}_0^{(1)}, \hat{a}_1^{(1)}$ при появлении следующего (t+1) члена ряда производится по формулам

$$\hat{a}_0^{(1)}(t+1;\lambda) = (1-\lambda^2)x(t+1) + \lambda^2\hat{a}_0^{(1)}(t) + \lambda^2\hat{a}_1^{(1)}(t)$$

$$\hat{a}_1^{(1)}(t+1;\lambda) = (1-\lambda^2)x(t+1) - (1-\lambda^2)\hat{a}_1^{(1)}(t) - (1-\lambda^2)\hat{a}_0^{(1)}(t) + \hat{a}_1^{(1)}(t) \quad (13)$$

Для случая квадратичного тренда, т.е. $Ex(t+\tau) = a_0 + a_1\tau + a_2\tau^2$

Выпишем для этого случая так называемые «формулы обновления», позволяющие пересчитывать оценки $\hat{a}_0^{(2)}(s;\lambda)$, $\hat{a}_1^{(2)}(s;\lambda)$, и $\hat{a}_2^{(2)}(s;\lambda)$, полученные как решения оптимизационной задачи (10) при *t=s*, когда появляется очередное *(s+1)*-е наблюдение временного ряда *x(s+1)*:

$$\hat{a}_0^{(2)}(s+1;\lambda) = \hat{a}_0^{(2)}(s;\lambda) + \hat{a}_1^{(2)}(s;\lambda) + \hat{a}_2^{(2)}(s;\lambda) + (1-\lambda^3)\hat{\varepsilon}(s+1)$$

$$\hat{a}_1^{(2)}(s+1;\lambda) = \hat{a}_1^{(2)}(s;\lambda) + 2\hat{a}_2^{(2)}(s;\lambda) + \frac{3}{2}(1-\lambda)(1-\lambda^2)\hat{\varepsilon}(s+1)$$

$$\hat{a}_2^{(2)}(s+1;\lambda) = \hat{a}_2^{(2)}(s;\lambda) + \frac{1}{2}(1-\lambda)^3\hat{\varepsilon}(s+1) \quad (14)$$

где $\hat{\varepsilon}(s+1) = x(s+1) - \hat{a}_0^{(2)}(s;\lambda) - \hat{a}_1^{(2)}(s;\lambda) - \hat{a}_2^{(2)}(s;\lambda)$ – ошибка прогноза на последнем шаге.

2.1.2 Метод Хольта

В этом методе была предпринята попытка ослабить ограничения метода Брауна, связанные с его однопараметричностью за счет введения двух параметров сглаживания λ_1, λ_2 $(0 < \lambda_1 < 1, 0 < \lambda_2 < 1)$. В его модели прогноз $\hat{x}(t;l)$ на l тактов времени в текущий момент t также определяется линейным трендом вида

$$\hat{x}(t;l) = \hat{a}_0(t;\lambda_1,\lambda_2) + l\hat{a}_1(t;\lambda_1,\lambda_2) \tag{15}$$

где обновление прогнозирующих коэффициентов проводится по формулам:

$$\begin{gathered}\hat{a}_0(t;\lambda_1,\lambda_2) = \lambda_1 x(t) + (1-\lambda_1)(\hat{a}_0(t;\lambda_1,\lambda_2) + \hat{a}_1(t;\lambda_1,\lambda_2)) \\ \hat{a}_1(t+1;\lambda_1,\lambda_2) = \lambda_2(\hat{a}_0(t;\lambda_1,\lambda_2) + \hat{a}_1(t;\lambda_1,\lambda_2)) - (1-\lambda_2)\hat{a}_1(t;\lambda_1,\lambda_2)\end{gathered} \tag{16}$$

Таким образом, прогноз по данному методу является функцией прошлых и текущих данных, параметров λ_1, λ_2, а также начальных значений $\hat{a}_1(t;\lambda_1,\lambda_2)$ и $\hat{a}_0(t;\lambda_1,\lambda_2)$.

2.1.3 Метод Хольта-Уинтерса

Метод Хольта был развит Уинтерсом так, чтобы он охватывал, помимо линейного тренда, еще и сезонные эффекты. Прогноз, сделанный в момент t на l тактов вперед, равен

$$\hat{x}(t;l) = [\hat{a}_0(t) + l\hat{a}_1(t)]\omega(t+l-T) \tag{17}$$

где $\omega(\tau)$ – так называемый коэффициент сезонности, а T – число временных тактов, содержащихся в полном сезонном цикле, т.е. в одном году (например, при месячных данных T=12).Сезонность в этой формуле представлена мультипликативно. Метод использует три параметра сглаживания (адаптации) $\lambda_1, \lambda_2, \lambda_3$ $(0 < \lambda_j < 1, j = 1,2,3)$, а его формулы пересчета обновления (прогнозирующих коэффициентов) имеют вид:

$$\hat{a}_0(t) = \lambda_1 \frac{x(t+1)}{\omega(t+1-T)} + (1-\lambda_1)\left[\hat{a}_0(t) + \hat{a}_1(t)\right]$$

$$\omega(t+1) = \lambda_2 \frac{x(t+1)}{\hat{a}_0(t+1)} + (1-\lambda_2)\omega(t+1-T)$$

$$\hat{a}_1(t+1) = \lambda_3\left[\hat{a}_0(t) + \hat{a}_1(t)\right] + (1-\lambda_3)\hat{a}_1(t) \tag{18}$$

Как и в методе Хольта, прогноз здесь строится на основании прошлых значений временного ряда, параметров адаптации $\lambda_1, \lambda_2, \lambda_3$, а также начальных значений $\hat{a}_0(0), \hat{a}_1(0), \omega(0)$.

2.1.4 Аддитивная модель сезонности Тейла-Вейджа

На практике часто встречаются экспоненциальные тенденции с мультипликативно наложенной сезонностью. Поэтому перед использованием аддитивной модели члены анализируемого временного ряда обычно заменяют их логарифмами и тем самым преобразуют экспоненциальную тенденцию в линейную и одновременно – мультипликативную сезонность на аддитивную. Преимущество же аддитивной модели – в относительной простоте ее вычислительной реализации. Будем считать, что логарифмическое преобразование исходных данных уже выполнено, и рассмотрим аддитивную модель вида

$$x(\tau) = a_0(\tau) + \omega(\tau) + \delta(\tau)$$

$$a_0(\tau) = a_0(\tau-1) + a_1(\tau) \tag{19}$$

где $a_0(\tau)$ – уровень процесса после элиминирования сезонных колебаний, $a_1(\tau)$ - аддитивный коэффициент роста, $\omega(\tau)$ - аддитивный коэффициент сезонности и $\delta(\tau)$ – белый шум.

Прогноз, сделанный в момент t на l временных тактов вперед, подсчитывается по формуле

$$\hat{x}(t;l) = \hat{a}_0(t) + l\hat{a}_1(t) + \omega(t + l - T) \qquad (20)$$

где коэффициенты $\hat{a}_0, \hat{a}_1$ и $\hat{\omega}$ вычисляются рекуррентным способом с помощью следующих формул обновления:

$$\begin{gathered}
\hat{a}_0(\tau) = \hat{a}_0(\tau - 1) + \hat{a}_1(\tau - 1) + \lambda_1[x(\tau) - \hat{x}(\tau - 1;1)] \\
\hat{a}_1(\tau) = \hat{a}_1(\tau - 1) + \lambda_1\lambda_2[x(\tau) - \hat{x}(\tau - 1;1)] \\
\hat{\omega}(\tau) = \hat{\omega}(\tau - T) + (1 - \lambda_1)\lambda_3[x(\tau) - \hat{x}(\tau - 1;1)]
\end{gathered} \qquad (21)$$

В этих соотношениях T – это число временных тактов, содержащихся в полном сезонном цикле (как правило, в году), а $\lambda_1, \lambda_2, \lambda_3$ – параметры адаптации.

Подбор подходящих значений для параметра адаптации является узким местом всех методов, основанных на экспоненциальном сглаживании. Обычно решение этой задачи осуществляется с помощью подбора подходящего значения на компьютере.