ИССЛЕДОВАНИЕ СВОЙСТВ КРИТЕРИЕВ ТРЕНДА И СЛУЧАЙНОСТИ МЕТОДАМИ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Щеглов А.Е., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: tleha@ngs.ru

Упорядоченная последовательность $\{y_t, t = 1,...,n\}$ наблюдений значений некоторой переменной, произведенных через равные промежутки времени, называется *временным рядом* [1]. Типичными примерами временных рядов являются наблюдения за курсом доллара в бизнесе и экономике, объем урожая в сельском хозяйстве, электрические сигналы в технических областях, температура воздуха в метеорологии.

Временные ряды могут складываться из [1]: тренда или систематической составляющей; сезонной составляющей; случайной составляющей (шума).

Существует множество критериев тренда и случайности, предназначенных для проверки гипотез о случайности расположения выборочных данных. При проверке отсутствия тренда (случайности) в математическом ожидании предполагается, что наблюдается временной ряд $y_1,..., y_n$ значений взаимно независимых случайных величин с математическими ожиданиями $\mu_1,...,\mu_n$ и одинаковыми (но неизвестными) дисперсиями. Проверяется нулевая гипотеза $H_0 : \mu_i = \mu,\ i = 1,2,...,n\}$, о том, что все выборочные значения принадлежат к одной генеральной совокупности со средним μ, против альтернативы о наличии тренда $H_1 : |\mu_{i+1} - \mu_i| > 0,\ i = 1,2,...,n-1$.

Цель работы заключалась в исследовании того, насколько хорошо распределения статистик некоторых критериев тренда и случайности согласуются со своими асимптотическим распределениями при различных законах случайных составляющих и различных объемах выборок. Исследования проводились с помощью методики компьютерного моделирования [2].

Для исследований выбраны следующие критерии [3].

Критерий Вальда-Волфовитца. Пусть $\tilde{y}$ – выборочная медиана ряда, значения $y_t \geq \tilde{y}$ обозначаются символом a, а значения $y_t < \tilde{y}$ – символом b. Статистикой критерия является общее число N_s полученных серий элементов a и b (под *серией* понимается последовательность элементов одной из нескольких выборок в упорядоченной по возрастанию объединенной выборке, ограниченная элементами другой выборки). Если n_a и n_b – соответственно количества элементов a и b в последовательности, то при $n_a, n_b > 20$ справедлива аппроксимация распределения статистики

$$N_s^* = \frac{N_s - (2n_a n_b / (n_a + n_b) + 1)}{\sqrt{2n_a n_b (2n_a n_b - n_a - n_b) / ((n_a + n_b)^2 (n_a + n_b - 1))}}$$

стандартным нормальным законом. При заданном уровне значимости α гипотеза H_0 отклоняется, если $\left|N_s^*\right| > u_{(1+\alpha)/2}$, где $u_{(1+\alpha)/2}$ - соответствующий квантиль стандартного нормального закона $N(0,1)$.

Критерий числа серий знаков первых разностей. Для выборки $y_1,...,y_n$ вычисляются $n-1$ значений z_i, равных -1, если $y_{i+1} < y_i$, равных $+1$, если $y_{i+1} > y_i$, и равных 0, если $y_{i+1} = y_i$. Количество серий Z в ряду z_i является статистикой критерия. При $n > 30$ распределение Z приближается нормальным законом с математическим ожиданием $M(Z) = (2n-1)/3$ и дисперсией $D(Z) = (16n-29)/90$. Гипотеза H_0 отклоняется при $\left|Z^*\right| = \frac{\left|Z - M(Z)\right|}{\sqrt{D(Z)}} \geq u_{(1+\alpha)/2}$.

Критерий инверсий. Если для ряда $y_1, y_2,..., y_n$ выполняется $y_i > y_j$, где $i+1 \leq j \leq n$, то имеет место инверсия. Общее число инверсий I является статистикой критерия. При $n \geq 20$ статистика I распределена приблизительно нормально с $M(I) = n(n-1)/4$ и $D(I) = (2n^3 + 3n^2 - 5n)/72$. Гипотеза о случайности отклоняется при $\left|I^*\right| = \frac{\left|I - M(I)\right|}{\sqrt{D(I)}} \geq u_{(1+\alpha)/2}$.

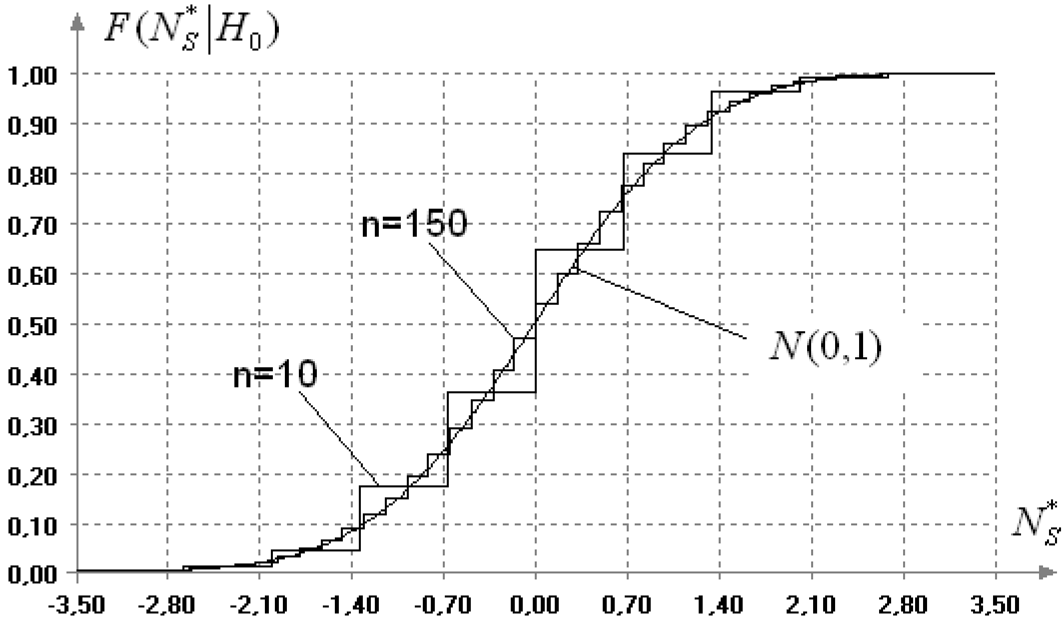

Рис. 1. Распределения статистики критерия Вальда-Волфовитца

Данные критерии являются непараметрическими, и распределения случайных составляющих не влияют на распределения статистик этих критериев, что подтверждается и результатами моделирования.

Исследование распределений статистик при разном объеме выборок показало, что областью определения статистик критериев Вальда-Волфовитца и числа серий знаков первых разностей является область дискретных значений. Даже при достаточно больших объемах выборок ($n = 100$) дискретные распределения статистик существенно отличаются от стандартного нормального закона. Лишь при $n \geq 150$ гипотеза о согласии не отвергается (достигаемый уровень значимости 0.1), но визуально ступенчатость графиков остается заметной (рис. 1). Поэтому использование стандартного нормального закона для определения достигнутого уровня значимости может приводить к ошибкам, особенно при небольших объемах выборок.

Распределение статистики критерия инверсий также является дискретным, но уже при объеме выборок $n = 30$ дискретность практически незаметна и гипотеза о согласии со стандартным нормальным законом не отвергается при достигнутом уровне значимости 0.19 (рис. 2). Таким образом, в данном случае при проверке гипотезы о случайности, начиная с $n \geq 30$, будет корректным использование квантилей стандартного нормального закона.

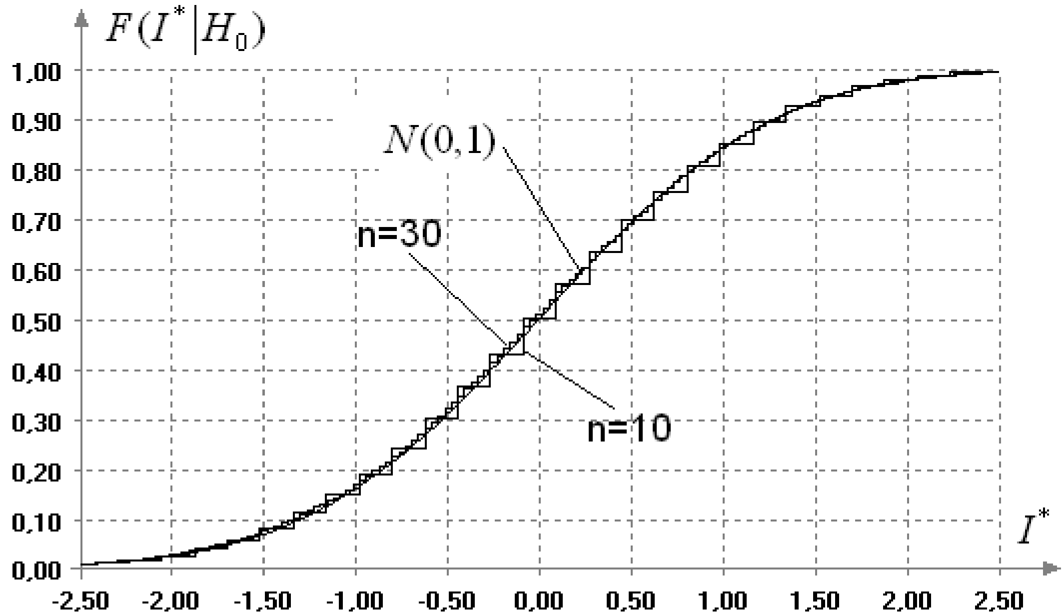

Рис. 2. Распределения статистики критерия инверсий

Работа выполнена при поддержке РФФИ (проект № 09-01-00056а), Аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы" (проект № 2.1.2/11855) и Федеральной целевой программы Минобрнауки РФ “Научные и научно-педагогические кадры инновационной России”.

Литература

1. Абденов А.Ж. Динамические эконометрические модели: Учебное пособие. – Новосибирск: Изд-во НГТУ, 2002. – 72 с.
2. Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учеб. пособие. – Новосибирск: Изд-во НГТУ, 2004. – 120 с.
3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.