СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ И ИССЛЕДОВАНИЕ СВОЙСТВ КРИТЕРИЕВ

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

Под *временным рядом* понимается упорядоченная последовательность $\{y_t, t = 1,\ldots,N\}$ наблюдений значений некоторой переменной, сделанных через равные промежутки времени [1]. Наблюдения за курсом доллара в бизнесе и экономике, объем урожая в сельском хозяйстве, электрические сигналы в технических областях, температура воздуха в метеорологии – все это примеры временных рядов [2].

Типичные временные ряды в экономике и в других областях могут складываться из [1]:

– тренда или систематической составляющей;

– сезонной составляющей;

– случайной составляющей (шума).

Временные ряды могут быть стационарными и нестационарными [2]. Временной ряд называется *стационарным*, если его вероятностные характеристики (математическое ожидание, дисперсия) постоянны. Временной ряд называется *нестационарным*, если хотя бы одна из вероятностных характеристик непостоянна. Многие часто встречающиеся на практике временные ряды являются нестационарными.

Два разных типа нестационарных по отношению к среднему временных рядов могут быть приведены к стационарному виду с помощью взятия последовательных разностей. Это ряды с детерминированным трендом (TS – trend stationary) и ряды, имеющие стохастический тренд (DS – difference stationary) [2, 3].

Принципиальное различие между этими двумя типами рядов выражается в том, что TS ряд также можно привести к стационарному виду с помощью выделения линейного тренда, тогда как вычитание детерминированной составляющей из DS ряда оставляет его нестационарным [3].

На практике TS и DS ряды можно спутать. Поэтому надо проводить четкое различие между данными рядами. Определение принадлежности рядов классам TS и DS важно для правильного построения долгосрочных регрессионных моделей [3].

Для решения вопроса об отнесении исследуемого ряда y_t к классу TS или DS рядов предложено множество статистических критериев, большинство из которых решает задачу в классе ARMA моделей (авторегрессионных со скользящим средним). В этом случае проверка нулевой гипотезы H_0 о том, что некоторый ARMA ряд является DS рядом, может быть сведена к проверке гипотезы о том, что авторегрессионный полином имеет хотя бы один корень, равный 1. При этом о нулевой гипотезе H_0 говорят как о *гипотезе единичного корня* [3].

Все эти критерии, как правило, основываются на предположении нормальности случайных составляющих. На практике же гипотеза о нормальности шума выполняется далеко не всегда. Поэтому цель настоящей работы заключалась в исследовании свойств некоторых критериев, используемых для различения TS и DS рядов, в зависимости от закона распределения шума.

Для исследований были выбраны два критерия Дики-Фуллера [3]. В обоих критериях рассматривается следующая статистическая модель в форме процесса авторегрессии:

$$y_t = a_1 y_{t-1} + \varepsilon_t,\ t = 1,..n. \tag{1}$$

Статистика первого рассмотренного критерия имеет следующий вид:

$$df = n(\bar{a}_1 - 1), \tag{2}$$

где $\bar{a}_1$ – оценка параметра a_1 по методу наименьших квадратов. В данном случае нулевая гипотеза $H_0:\ a_1 = 1$ (ряд является DS рядом), а альтернативная гипотеза – $H_1:\ a_1 < 1$ (ряд является TS рядом).

Таблицы критических значений распределения статистики (2) были построены Фуллером в предположении, что случайные составляющие распределены по нормальному закону.

Статистика второго рассмотренного критерия имеет следующий вид:

$$t = \frac{\bar{a}_1 - 1}{s(\bar{a}_1)}, \tag{3}$$

где $\bar{a}_1$ – также оценка параметра a_1 по методу наименьших квадратов, а $s(\bar{a}_1)$ – оцененная стандартная ошибка $\bar{a}_1$.

В настоящей работе с помощью методики компьютерного моделирования смоделировано распределение статистик (2) и (3) при разных объемах выборки n и случайных составляющих, имеющих двустороннее экспоненциальное распределение со значениями параметра формы $\lambda = 2$ (нормальное распределение), $\lambda = 1$ (распределение Лапласа), $\lambda = 4, \lambda = 0.5$ и $\lambda = 0.25$.

При нормальном распределении случайных составляющих критические значения получились достаточно близкими к критическим значениям из таблиц, построенных Фуллером.

В случаях, когда шум имеет отличное от нормального распределение, распределения статистик (2) и (3) отличаются от распределений при нормальном шуме.

В таблицах 1 и 2 приведены полученные критические значения распределений статистик (2) и (3), соответственно, при 5% уровне значимости для некоторых объемов выборки при разном распределении случайных составляющих. На рисунке 1 приведено эмпирическое распределение статистики (2) при $n = 25$ и разном распределении шума.

Таблица 1. Критические значения распределения статистики (2)

n	$\lambda = 4$	$\lambda = 2$	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0.25$
25	-7.32302	-7.35309	-7.16105	-6.85325	-6.45845
50	-7.91586	-7.6952	-7.60992	-7.40507	-6.93471
100	-7.97669	-7.851	-7.78701	-7.63853	-7.26963

Таблица 2. Критические значения распределения статистики (3)

n	$\lambda = 4$	$\lambda = 2$	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0.25$
25	-1.96595	-1.94573	-1.91629	-1.85177	-1.74865
50	-1.95261	-1.94775	-1.93187	-1.8981	-1.84567
100	-1.93553	-1.93384	-1.9334	-1.88523	-1.80128

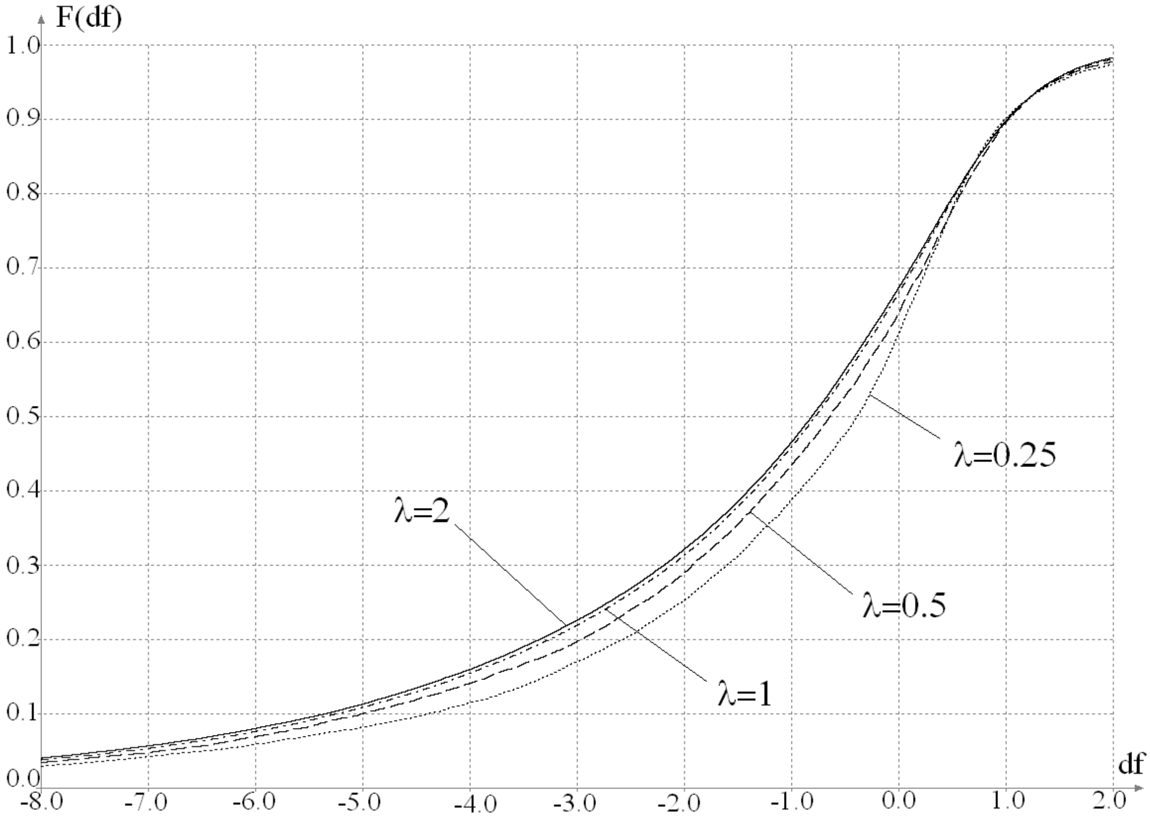

Рис. 1. Эмпирическое распределение статистики (2)

Настоящие исследования выполнены при частичной поддержке РФФИ (проект № 09-01-00056-а), Федерального агентства по образованию в рамках АВЦП “Развитие научного потенциала высшей школы” и ФЦП “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. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования: учеб. пособие. – Новосибирск: Изд-во НГТУ, 2008. – 272 с.
3. Носко В. П. Эконометрика. Введение в регрессионный анализ временных рядов. – Москва, 2002.