

МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ И ИССЛЕДОВАНИЕ СВОЙСТВ ОЦЕНОК И СТАТИСТИК В ЗАВИСИМОСТИ ОТ ВЕРоятНОСТНЫХ ХАРАКТЕРИСТИК ШУМА¹

Краснер И.А., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: krilar@yandex.ru

Во многих статистических задачах обрабатываемые данные представляют собой временной ряд наблюдений, регистрируемых последовательно во времени. Наблюдения такого рода преобладают в некоторых геофизических науках (метеорология, сейсмология) и в социальных науках, особенно в экономике. Они играют важную роль в медицинских исследованиях, в задачах анализа траекторных наблюдений и т. д.

Временным рядом называют последовательность наблюдений, упорядоченную по времени. Основной чертой, выделяющей анализ временных рядов, является существенность порядка, в котором производятся наблюдения.

Цели анализа временных рядов могут быть различными. Можно стремиться предсказать будущее на основании знания прошлого или пытаться выяснить механизм, лежащий в основе процесса, и управлять им.

Рассматривались следующие модели

$$x_i = \sum_{t=1}^p A_t x_{t-1} + e_i;$$

$$x_i = \sum_{t=1}^p A_t x_{t-1} + \sum_{k=1}^q B_k e_{k-1};$$

$$x_i = \sum_{k=1}^q B_k e_{k-1},$$

где

x_i – значения ряда в момент времени t ;

A_t – коэффициенты при лаговых значениях ряда;

B_k – коэффициенты при значениях шума;

e_i – случайный шум.

Все исследуемые ряды являлись стационарными.

Обычно случайный шум считают гауссовским белым шумом. Но часто условия задачи отличаются от классических постановок. Например, шум может быть распределен не по нормальному закону или значения шума в разные моменты времени могут быть коррелированы. В этом случае нельзя делать вы-

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 06-01-00059-а)

воды о некоторых свойствах оценок, например о виде закона распределения, опираясь на классическую теорию. В то же время, моделируя ситуацию и исследуя результаты моделирования, можно получить дополнительную информацию о модели и ее свойствах.

В данной работе при исследовании закономерностей, связанных с вышеприведенными моделями, рассматривался некоррелированный, отличный от нормального шум.

В качестве законов распределения шума рассматривались распределение Лапласа и равномерное распределение

В соответствии с целями исследований разработано программное обеспечение, позволяющее моделировать многомерные временные ряды произвольной размерности. С помощью данного программного обеспечения проведено моделирование временных рядов и получены выборки оценок коэффициентов при лаговых значениях ряда. Для анализа этих выборок (проверка согласия с заданным распределением, выбор наилучшего распределения) использовалась программа ISW “Статистический анализ одномерных наблюдений ” версия: 4.2.0.11

Исследования показали, что при отклонении закона распределения шума от нормального, распределения оценок параметров все же можно достаточно хорошо описать с помощью нормального закона.

Проведено исследование распределений статистик следующих критериев проверки на наличие тренда и корреляций в наблюдаемом временном ряде:

- критерия Манна-Кендалла проверки на наличие тренда;
- коэффициента ранговой корреляции Спирмена

$$\eta = 1 - \frac{6}{T(T^2 - 1)} \sum_{t=1}^T (\theta_t - t)^2; \quad (1)$$

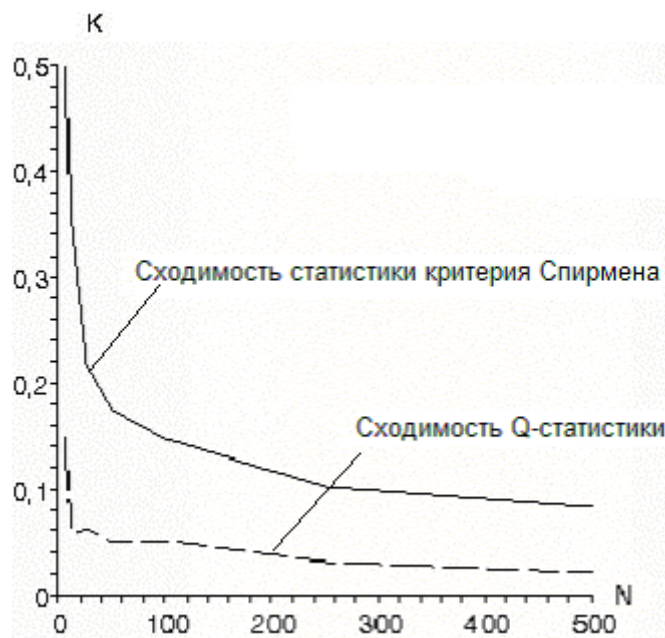
- Q – статистики Льюнга – Бокса

$$Q(r) = T(T + 2) \sum_{k=1}^m \frac{r_k^2}{T - k}. \quad (2)$$

Эти критерии могут применяться не только к значениям ряда, но и к остаткам модели. Например, после вычитания из ряда выявленного тренда получаются остатки, которые можно расценивать, как оценки случайной компоненты. Наличие в остатках каких-либо зависимостей свидетельствует о том, что модель неполна или некорректна.

Для проверки согласия эмпирических распределений статистик (1) и (2) исследуемых критериев с теоретическими использовался критерий Колмогорова. Как и предполагалось, эмпирические распределения рассматриваемых статистик сходятся к соответствующим теоретическим распределениям независимо от параметров шума, хотя скорость сходимости зависит от рассматриваемого закона распределения шума. Например, чем больше дисперсия распределения, соответствующего шуму, тем медленнее распределение Q – статистики сходится к “классическому” теоретическому.

Результаты исследований показали, что распределение статистики Льюнга–Бокса лучше сходится к своему теоретическому распределению (какому), чем распределение рангового коэффициента корреляции Спирмена. Результаты сходимости иллюстрирует следующий рисунок, на котором показано уменьшение расстояния Колмогорова K между эмпирическим и соответствующим теоретическим распределениями статистик в зависимости от количества наблюдений N .



По поводу оценок параметров A_t временного ряда можно отметить следующее:

1. Дисперсия оценок параметров при больших выборках скорее зависит от дисперсии ошибок, чем от вида их распределения ошибок.
2. Закон распределения ошибок влияет на сходимость распределений оценок параметров к нормальному закону (распределение оценок сходятся к нормальному при больших объемах выборок).
3. Распределения оценок параметров, значения которых близки или равны нулю, сходятся к нормальному закону существенно медленнее.
4. В случае относительно небольших выборок дисперсии оценок зависят как от свойств шума (дисперсии, закона распределения), так и от самой модели.

Так как исследования проводились на конкретных моделях, с конкретными свойствами шумов, все выводы в настоящей работе могут касаться только этих моделей и пока не имеют общего характера.

1. EPA QA/G-9S. Data Quality Assessment: Statistical Methods for Practitioners. <http://www.epa.gov/quality/qs-docs/g9s-final.pdf>
2. Т. Андерсон. Статистический анализ временных рядов. – М.: Мир, 1976. – 755 с.
3. Helmut Lutkepohl. New Introduction to Multiple Time Series Analysis Springer 2005. 764p. ISBN 3-540-40172-5