

# Сравнительный анализ алгоритмов статистического прогнозирования

Трофимович А.Ф.  
Кафедра ИТАС, ФИТУ  
БГУИР  
Минск, Беларусь  
e-mail: alex\_tr@tut.by

**Аннотация** – В работе проводится анализ существующих алгоритмов статистического прогнозирования. Предлагаются варианты сравнения эффективности различных алгоритмов прогнозирования.

**Ключевые слова:** прогнозирование; статистика; алгоритм; сравнение

## I. ВВЕДЕНИЕ

Существует достаточно много исследований в области прогнозирования. В настоящее время существует более 100 методов прогнозирования, начиная с общенаучных (анализ и синтез, экстраполяция и интерполяция, индукция и дедукция, аналогия, гипотеза, эксперимент и т.д.) и заканчивая частнонаучными. Однако наиболее распространенных есть 10-15 обще- и межнаучных методов. К ним относятся:

- *экстраполяция* (с учётом особенностей динамики развития объекта, возможных отклонений временного ряда под воздействием факторов прогностического фона);
- *моделирование* (имитационное, игровое, операциональное, сетевое и др. модели);
- *опрос экспертов и населения;*
- *историческая аналогия;*
- *прогнозные сценарии;*
- *матрицы взаимовлияющих факторов* типа "проблемы - возможные способы их решения", "затраты - выпуск" и т.п., а также методы, основанные на *построении графов* и "дерева проблем" или "дерева целей", методы, основанные на использовании патентов и т.д.

## II. МОДЕЛИ ПРОГНОЗИРОВАНИЯ

В случае, когда прогнозируемый параметр представлен в количественной шкале, рассматривались модели:

- *линейная регрессионная модель (LRM);*
- *общая регрессионная модель (GRM);*
- *модель корректировки отклонений (ECM);*
- *векторная авторегрессионная модель (VAR);*
- *векторная модель корректировки отклонений (VECM);*
- *авторегрессионная модель интегрированного скользящего среднего (ARIMA);*
- *нейросетевые модели (NSM).*

Рассматривались следующие методы оценивания:

- *метод наименьших квадратов (OLS),*
- *общий метод наименьших квадратов (GLS),*
- *метод взвешенных наименьших квадратов (WOLS),*
- *двухшаговый метод наименьших квадратов (TSLS),*

В настоящее время широкой популярностью пользуется метод "нейросетевого анализа". По своей

сути это использование правил и законов работы нервных клеток головного мозга человека. Особенностью таких программ является то, что они могут обучаться.

Нейросетевые принципы уже помимо программного, реализованы и на аппаратном уровне, так набирает темп индустрия нейрокомпьютеров с гигантской быстротой вычислений. Нейросетевой анализ можно использовать для решения задач:

- *аппроксимации и прогнозирования;*
- *многомерной классификации.*

В первом случае решается задача подобная регрессионному анализу и прогнозированию временных рядов, во втором, – кластерному анализу.

Однако нейросетевой анализ не делает никакого исходного предположения о форме связи между факторами и зависимой переменной, у нее нет четкого алгоритма классификации. По сути нейросетевая модель подбирает форму связи наилучшим образом описывающую обучающую выборку.

Известные преобразования многих моделей трендов путём логарифмирования или перехода к обратным величинам, сводящие их к линейным по параметрам, требуют «искусственных» предположений о виде стохастической компоненты. Например, при моделировании ряда динамики экспоненциальной функцией стохастическую компоненту для удобства логарифмирования всей модели предполагают в виде показателя мультипликативной по отношению к тренду экспоненты с логнормальной функцией распределения вероятностей значений. Определение параметров моделей (обычно путём применения метода наименьших квадратов (МНК)), над преобразованными величинами, а не над исходными, даёт погрешность, особенно при осуществлении прогнозирования.

Логарифмирование будет неработоспособно при аддитивной стохастической компоненты, которую большинство исследователей считают более естественным, а также при присутствии в структуре ряда динамики аддитивной или мультипликативной знакопеременной сезонной (или циклической) компоненты.

Моделирование многокомпонентных структур традиционно осуществляется путем реализации двух последовательных этапов: моделирования и выделения тренда (при этом исходная выборка для моделирования тренда искажена присутствием, например, сезонной компоненты), а затем – параметризацией сезонной компоненты по новой выборке. При этом погрешность моделирования на первом этапе присутствует в новой выборке для реализации второго этапа, увеличивая общую погрешность.

До настоящего времени практически не моделируются (не рассматриваются при идентификации) случаи присутствия в рядах динамики (кроме известного случая линейной регрессии) временного тренда и экзогенных воздействий, что

также ведёт к погрешностям, сужает область реального эконометрического моделирования и прогнозирования;

Зачастую предполагают априори известными значения некоторых параметров моделей (например, в моделях логистической динамики), что или ограничивает область применения моделирования и прогнозирования, или ведёт к погрешности при неточном знании этих параметров.

Наконец, отсутствует общий математический подход в рамках линейных уравнений к решению задачи моделирования и прогнозирования широкого класса моделей, а известны лишь отдельные эвристические решения по методам моделирования. Применение нелинейного МНК нежелательно даже не в силу их относительной сложности, но скорее из-за необходимости исследования сходимости методов решения, обоснования выбора начальной точки и главное, вследствие проблем, возникающих при нелинейном методе наименьших квадратов со смещенностью и состоятельностью оценок параметров, особенно если стохастическая компонента обладает свойством гетероскедастичности [1].

Авторегрессии известны уже несколько десятков лет, но в них никоим образом, в виду отсутствия механизма их конструирования, не определялась связь между порядком, оценками коэффициентов  $\lambda$  авторегрессии и типами, параметрами моделей моделируемого ряда динамики. Позволяя осуществлять прогноз уровней (значений) ряда динамики, фиксировать его «разладку», непараметрические авторегрессии не отвечали на вопрос о том, что было её причиной – смена параметров модели или смена её класса. Развитие регрессионного подхода в последние годы шло в направлении более сложных и специализированных для длинных рядов динамики, но также непараметрических, моделей: ARMA, ARIMA, ARCH, GARCH, IGARCH, EGARCH и др.

В настоящее время существуют и другие методы решения задачи прогнозирования. В частности, на основе приведенного в [2] подхода был получен алгоритм прогнозирования реализации векторного случайного процесса по наблюдениям без ошибок на основе общей теории статистических решений с привлечением многомерно-матричного математического подхода [3]. Указанный алгоритм показал неплохие результаты практического использования при прогнозировании количественных характеристик погоды (температуры воздуха и атмосферного давления) [4]. Помимо линейных алгоритмов статистического прогнозирования существуют и нелинейные [5], обладающие как достоинствами (в отдельных случаях повышается точность прогнозирования), так и недостатками (возрастает вычислительная сложность).

### III. ЗАКЛЮЧЕНИЕ

Сравнительный анализ эффективности алгоритмов проводился с использованием средней квадратичной ошибки прогнозирования, которая может быть вычислена для заданных моделей прогнозируемых случайных последовательностей: стационарных случайных процессов 2-го порядка, гауссовских временных рядов, временных рядов с трендом. Достаточно сложно проверить эффективность всех известных алгоритмов статистического прогнозирования, показать все возможные направления, методы и пути решения проблемы прогнозирования. Эта проблема многослойная (многоуровневая) с множеством прямых и обратных связей, и, как правило, с нелинейными зависимостями, с множеством факторов, которые невозможно либо учесть, либо они не известны.

- [1] Клейнер Г.Б., Смоляк С.А. Эконометрические зависимости: принципы и методы построения. - М.: 2000. - 104 с.
- [2] Колмогоров, А.Н. Интерполирование и экстраполирование стационарных случайных последовательностей / А.Н. Колмогоров // Известия АН СССР. Сер. матем. – 1941. – Т. 5. – С. 3–14.
- [3] Муха, В.С. Статистическое векторное прогнозирование количественных характеристик погоды / В.С. Муха // Информационные системы и технологии (IST'2004). Материалы междунар. конф., Минск, Беларусь, 8 – 10 ноября 2004. Ч. 2. – Минск, 2004. – С. 195 – 200.
- [4] Муха, В.С. Об эффективности скалярного статистического прогнозирования метеорологических процессов / В.С. Муха, А.Ф. Трофимович // Информатика – № 1 (17). – 2008. – С. 113 – 124.
- [5] Fan, J. and Yao, Q. (2003). Nonlinear time series. Nonparametric and parametric methods. Springer, New York)
- [6] Андерсон, Т. Статистический анализ временных рядов / Т. Андерсон – М.: Мир, 1976. – 756 с.
- [7] Айвазян С.А., Мхитарян В.С. Прикладная статистика. Теория вероятностей и прикладная статистика. - М.: ЮНИТИ – ДАНА. 2001. - 656 с.
- [8] Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып.2. - М.: Мир. 1974. - 197 с.
- [9] Соколов Н.П. Введение в теорию многомерных матриц. – Киев: Наукова думка, 1972. – 175 с.
- [10] М. Де Гроот. Оптимальные статистические решения. – М.: Мир, 1974. – 496 с.
- [11] Рао С.Р. Линейные статистические методы и их применение. – М.: Наука, 1968. – 548 с.
- [12] Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. Пер. с английского. - М.: Мир, 1982. - 488с.
- [13] Норман Дрейпер, Гарри Смит. Прикладной регрессионный анализ. Множественная регрессия (Applied Regression Analysis). – 3-е изд. – М.: «Диалектика», 2007. – С. 912.
- [14] Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980, с. 456.