УДК: 60.608

Мокренко Н.В.,
студент
3 курс, факультет «Информатика и системы управления»
Московский государственный технический университет им. Н.Э. Баумана
Россия, г. Москва
Головин И.В.,
студент магистратуры
2 курс, факультет «Информатика и системы управления»
Московский государственный технический университет им. Н.Э. Баумана
Россия, г. Москва
Научные руководители: Тоноян С.А.,
доцент кафедры «Системы обработки информации и управления»
Селиверстова А.В.,
аспирант кафедры «Системы обработки информации и управления»

ОБЗОР АВТОРЕГРЕССИОННЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Аннотация: *В данной статье проводится обзор и сравнение различных моделей авторегрессии, объясняется понятие авторегрессии, разбирается прогнозирование и анализ временных рядов, анализируется эффективность прогноза при помощи моделей авторегрессии.*

Ключевые слова: *авторегрессионная модель, AR, ARMA, ARIMA, ARCH, анализ, прогнозирование, временные ряды.*

Annotation: *This article reviews and compares various models of autoregression, explains the concept of autoregression, analyzes forecasting and analysis of time series, analyzes the effectiveness of forecasts using autoregression models.*

Keywords: *autoregressive model, AR, ARMA, ARIMA, ARCH, analysis, forecasting, time series.*

Введение

Цель исследования временного ряда – подбор модели, описывающей временной ряд и применяемый с целью прогноза его будущих значений.

Один из способов исследовать временной ряд и спрогнозировать его будущие значения – это построить на нем модель авторегрессии и скользящего среднего. Такая модель особенно эффективна для описания процессов, совершающих однородные колебания вокруг своего среднего значения. Как следствие, такие модели подходят лишь для стационарных рядов, среднее, дисперсия и автокорреляция которых лишь незначительно изменяются во времени и приблизительно постоянны.

Анализ временного ряда

Временным рядом называется упорядоченная во времени совокупность измерений одной из характеристик исследуемого объекта x_t, где t – порядковый номер анализируемого периода. Последовательные наблюдения временного ряда обладают признаком взаимозависимости, особенно к близко расположенным показателям. Информативность показателей ряда различна и зависит от удаленности от текущего момента (по мере удаления от текущего момента времени информационная ценность снижается). Достоверность статистических характеристик не будет увеличиваться пропорционально росту числу наблюдений.

Исследование временного ряда предполагает, что в данных содержатся две основные составляющие:

- систематическая составляющая (регулярная);
- случайный шум (ошибка), который может затруднять обнаружение регулярных компонент.

К регулярным составляющим временных рядов относится два класса компонент: это либо тренд, либо сезонная составляющая.

Тренд представляет собой общую систематическую линейную или нелинейную компоненту, которая может изменяться во времени. Сезонная составляющая – это периодически повторяющаяся компонента. Обе эти компоненты могут присутствовать в ряде одновременно.

Модель Бокса – Дженкинса

ARMA (модель авторегрессии и скользящего среднего) модель, также называется моделью Бокса – Дженкинса. Преимуществом модели считается формализованная и наиболее подробно разработанная методика, следуя которой можно подобрать параметры, наиболее подходящие к каждому конкретному временному ряду.

Общий вид модели ARMA:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

где Θ_j – параметр скользящего среднего, ε_t – белый шум, ϕ_i – параметр авторегрессии, c – константа.

Выбор параметров является основной задачей в модели ARMA(p, q), где p – порядок авторегрессионной части, q – порядок скользящей средней.

Перед построением авторегрессионных моделей необходимо провести тест Дики – Фуллера на стационарность ряда. Ряд называется строго стационарным (или стационарным в узком смысле), если его свойства не изменяются при изменении начала отсчета времени. В случае, если ряд проходит этот тест, можно строить на нем ARMA модель.

Выбрать наиболее оптимальную модель ARMA(p, q) можно на основании информационного критерия AIC (An Information Criterion, позднее именуется как Akiake Information Creterion), который был предложен в 1973 году:

$$AIC = -2\log(L) + 2(p + q + k + 1)$$

где L – функция правдоподобия, k = 1 при наличии константы и k = 0 при ее отсутствии.

Чем меньше будет значение AIC, тем лучше модель будет описывать имеющиеся данные.

Таким образом, перебирая параметры p и q, можно выбрать наиболее оптимальную авторегрессионную модель.

Модель ARIMA

ARIMA(p, d, q) является дополнением модели ARMA(p, q) для нестационарных временных рядов, которые можно свести к стационарным при помощи взятия разностей d-го порядка. Эти модели используются при работе с временными рядами для более глубокого понимания данных или более эффективного прогнозирования будущих значений ряда. Общий вид модели:

$$\Delta^d Y_t = \sum_{i=1}^{p} \alpha_i \Delta^d Y_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t$$

где Δ^d – оператор разности d-го порядка.

Модель SARIMA

Модель SARIMA(*p, d, q*)(*P, D, Q*)*s* является обобщением ARIMA-модели для временных рядов, в которых имеется ярко выраженная сезонная компонента. В такой модели помимо порядка авторегрессионной части, порядка скользящей средней и порядка разностей вводятся сезонные параметры (*P, D, Q, s*), позволяющие учитывать циклические колебания процесса при построении прогноза.

Построение модели можно осуществлять лишь в том случае, если ряд не проявляет нестационарность из сезона в сезон (например, тенденции к росту или снижению). Для проверки стационарности существуют различные тесты. Одним из таких является тест Канова – Хансен, который проверяет наличие тренда в сезонных компонентах.

Модель ARCH

Такие классические авторегрессионные модели, как ARMA, не всегда могут верно учитывать характеристики, которыми обладают финансовые временные ряды. Поэтому, необходимо расширение этих моделей.

Сама модель используется для исследования временных рядов, у которых относительный (по предыдущим значениям ряда) разброс значений ряда находится в зависимости от предшествующих значений дисперсий данного ряда и прочих условий.

Допустим, у нас есть временной ряд $X(t)$ некоторых экономических характеристик (стоимостей, индексов, ставок по кредитам и др.).

Тогда, преобразовав ряд $X(t)$ в ряд $Y(t) = \ln\frac{X(t+1)}{X(t)}$, получим непосредственно ряд индексов цен.

Пусть ряд имеет следующий вид:

$$u_t = \varepsilon_t \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2}$$

Тогда общий вид ARCH(q) модели:

$$\sigma_t^2 = V(u_t | u_{t-1}, \dots, u_{t-q}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2$$

где σ_t^2 – это условная дисперсия временного ряда.

Если ряд стационарный, то его безусловная дисперсия будет постоянна и равна:

$$\sigma^2 = \frac{\sigma_\varepsilon^2}{1 - \sum_{i=1}^{q} \alpha_i}$$

Оценка параметров данной модели может быть произведена методом наименьших квадратов.

Впервые ARCH модели были разработаны американским экономистом Робертом Инглом в 1982 году. Уже в 1986 году его датский коллега, Боллерслев, предложил обобщение этих моделей (GARCH).

Модель GARCH

GARCH является обобщением модели ARCH для временных рядов, условная дисперсия которых зависит от своих предыдущих значений.

В этом случае GARCH(p, q) модель (где p – порядок GARCH-членов σ^2 и q – порядок ARCH-членов u^2) описывается следующим образом:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

Необходимое условие стационарности ряда:

$$\sum_{j=1}^{p} \beta_j + \sum_{i=1}^{q} \alpha_i < 1$$

Безусловная дисперсия стационарного ряда равна:

$$\sigma^2 = \frac{\sigma_\varepsilon^2}{1 - \sum_{j=1}^{p} \beta_j - \sum_{i=1}^{q} \alpha_i}$$

Асимметричные модели GARCH

Асимметричные модели служат для того, чтобы при построении экономического прогноза учитывать асимметрию, возникающую время от времени на финансовом рынке. Отрицательные шоки чаще всего оказывают большее влияние на волатильность, чем положительные шоки, то есть изменчивость цены выше на падающем рынке, чем на растущем. Этот эффект называется эффектом левериджа (рычага). В рамках традиционных GARCH-моделей его объяснить невозможно, так как условная дисперсия зависит от квадратов прошлых значений ряда и не зависит от знаков.

Пример такой модели – QGARCH (квадратичная GARCH), предложенная Сентана в 1995 году:

$$\sigma_t^2 = \sigma^2 + a^T x_{t-q} + x_{t-q}^T A x_{t-q} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 , \qquad x_{t-q} = \left(u_{t-1}, \dots, u_{t-q}\right)^T$$

где A – симметрическая положительно определённая матрица, a – положительный вектор.

Эта модель предусматривает помимо эффекта левериджа также и вероятную связь влияния лагов благодаря внедиагональным элементам матрицы A. В случае, если матрица A является диагональной, а вектор a – нулевой, получатся стандартные модели GARCH. Если при диагональной матрице A вектор a не равен нулю, то получатся асимметричные GARCH.

Выводы

Модель авторегрессии – скользящего среднего является довольно простой, так как суть ее работы понятна и наглядна; однако, с другой стороны, она сложна тем, что в ней присутствует множество нюансов, требующих тщательного труда разработчика. Основным нюансом является, вне всякого сомнения, выбор наиболее эффективной модели для данного временного ряда из множества существующих.

Такие модели, в силу своей специфики, лучше всего подходят для краткосрочного прогнозирования. В зависимости от параметров исходного временного ряда и его стационарности выбирается наиболее подходящая модель. Однако, нельзя забывать золотое правило построения моделей прогнозирования

– сложные, статистически обоснованные модели необязательно дают более точные прогнозы, чем простые.

Заключение

В данной статье были рассмотрены наиболее популярные модели авторегрессии, объяснены различия между этими моделями, разобрана методология прогнозирования временных рядов при помощи этих моделей.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Лукашин, Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: Учеб. пособие. – М.: Финансы и статистика, 2003. – 416 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление: Пер. с англ. // Под ред. В.Ф. Писаренко. – М.: Мир, 1974, кн. 1. – 406 с.
3. Akaike, Hirotugu A new look at the statistical model identification. // Transactions on Automatic Control.1974, 19(6). P. 716–723.
4. Canova Fabio, Hansen Bruce E. Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability // Journal of Business & Economic Statistics. 1995. Vol. 13. P. 237-252.
5. Makridakis S., Hibon M. The M3 – competition: Results, conclusions and implications // International Journal of Forecasting. 2000. № 16. P. 451–476.
6. Генеральная совокупность: Энциклопедия экономики. [Электронный ресурс]. URL: https://economy-ru.info/info/19313/ (дата обращения: 06.03.2019).
7. Анализ временных рядов: Электронный учебник по статистике. [Электронный ресурс]. URL: http://statsoft.ru/home/textbook/modules/sttimser.html (дата обращения: 06.03.2019).