

Подбор стационарной модели ARMA для ряда наблюдений

$$ARMA(p, q) : x_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=0}^q \beta_j \varepsilon_{t-j}, \quad \beta_0 = 1, \quad \alpha_p \neq 0, \quad \beta_q \neq 0, \quad \varepsilon_t - \text{белый шум.}$$

Задача: При предположении, что некоторый временной ряд $x_1, x_2, \dots, x_T$ порождается моделью ARMA, необходимо подобрать конкурентную модель из этого класса моделей.

Решение этой задачи предусматривает три этапа:

1. **идентификация** модели;
2. **оценивание** модели;
3. **диагностика** модели.

На этапе идентификации производится выбор некоторой частной модели из всего класса ARMA, т.е. выбор значений p и q . Используемые при этом процедуры являются не вполне точными, что может при последующем анализе привести к выводу о непригодности идентифицированной модели и необходимости замены ее альтернативной моделью. На этом же этапе делаются предварительные грубые оценки коэффициентов $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$ идентифицированной модели.

На втором этапе производится вычисление оценок коэффициентов модели с использованием эффективных статистических методов. Для оцененных коэффициентов вычисляются приближенные стандартные ошибки, дающие возможность, при дополнительных предположениях о распределениях случайных величин $X_1, X_2, \dots$, строить доверительные интервалы для коэффициентов и проверять гипотезы об их истинных значениях с целью уточнения спецификации модели.

На третьем этапе применяются различные диагностические процедуры проверки адекватности выбранной модели имеющимся данным. Неадекватности, обнаруженные в процессе такой проверки, могут указать на необходимую корректировку модели, после чего производится новый цикл подбора, и т.д. до тех пор, пока не будет получена удовлетворительная модель.

В случае, если мы имеем дело с ситуацией, когда уже имеется достаточно отработанная и разумно интерпретируемая модель эволюции того или иного показателя, можно обойтись и без идентификации.

Идентификация стационарной модели ARMA

Основной отправной точкой для идентификации стационарной модели ARMA является различие поведения **автокорреляционных** и **частных автокорреляционных** функций временных рядов, соответствующих различным моделям ARMA. Ранее было рассмотрено поведение автокорреляционных функций для различных моделей ARMA. В частности, было показано, что:

1. автокорреляционные функции ρ_k процесса $AR(p)$, начиная с некоторого $k \geq p$, являются экспоненциально затухающими;
2. автокорреляционные функции ρ_k процесса $MA(q)$, начиная с некоторого $k \geq q$, обрываются ($\rho_k = 0$).

Ясно, что порядок авторегрессионной модели имеет большее значение, однако по поведению только автокорреляционной функции трудно идентифицировать даже порядок чистого (без MA составляющей) процесса авторегрессии. Решению этого вопроса помогает рассмотрение поведения частной автокорреляционной функции.

В определение автокорреляционной функции

$$\rho_\tau = \frac{1}{Var(X_t)} E\{(X_t - \mu)(X_{t-\tau} - \mu)\}$$

входит ковариация между значениями процесса, отстоящими на τ шагов о времени друг от друга. Однако, на поведение процесса $AR(p)$ статистически влияет не только его значение в момент, отстоящий на τ единиц назад, но и все промежуточные значения процесса между моментами t и $t - \tau$.

В этом случае возникает вопрос: какова линейная статистическая зависимость между значениями процесса в эти моменты, если мы устраним влияние всех промежуточных значений?

Коэффициент корреляции при исключении промежуточных значений называется **частным коэффициентом корреляции**.

Обозначим через φ_{kk} - k -ое значение частной автокорреляционной функции. По определению φ_{kk} - это коэффициент корреляции между x_{t-k} и x_t за вычетом той части x_t , которая линейно объяснена промежуточными шагами $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$.

Более строго, нас интересует коэффициент корреляции

$$Corr(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \dots - \varphi_{kk-1}x_{t-k+1}),$$

где $\varphi_{k1}, \varphi_{k2}, \dots, \varphi_{kk-1}$ - коэффициенты линейной комбинации, обеспечивающей минимальную среднеквадратичную ошибку предсказания

$$\min E\{(x_t - \varphi_{k1}x_{t-1} - \varphi_{k2}x_{t-2} - \dots - \varphi_{kk-1}x_{t-k+1})^2\}.$$

Если исследуемый процесс принадлежит к виду $AR(p)$, то $\varphi_{pp} = \alpha_p$. Более того, для такого процесса текущее значение выражается через ровно p предыдущих значений, и учет более ранних значений уже не может улучшить прогноз.

Это означает, что для процесса $AR(p)$ $\varphi_{kk} = 0$ при $k > p$.

Рассмотрим несколько простых примеров расчета значений частной автокорреляционной функции.

Процесс $AR(1)$. $X_t = \alpha X_{t-1} + \varepsilon_t$.

1. Рассмотрим регрессию: $X_t = \varphi_{11}X_{t-1} + \varepsilon_t$.

Умножаем уравнение на X_{t-1} , берем математическое ожидание от обеих частей и делим результат на автокорреляционную функцию $\gamma(0) = Cov(X_t, X_t)$.

Получаем:

$$\gamma(0) = Cov(X_t, X_t) = E[(X_t - E(X_t))(X_t - E(X_t))] = E(X_t, X_t), \text{ так как } E(X_t) = 0$$

$$\frac{E(X_t, X_{t-1})}{\gamma(0)} = \varphi_{11} \underbrace{\frac{E(X_{t-1}^2)}{\gamma(0)}}_{=1} + \frac{E(X_{t-1}, \varepsilon_t)}{\gamma(0)} \Rightarrow \rho_1 = \varphi_{11} = \alpha$$

2. Чтобы получить φ_{22} , построим теоретическую регрессию

$X_t = \varphi_{21}X_{t-1} + \varphi_{22}X_{t-2} + \varepsilon_t$. Как и раньше, умножаем обе части на X_{t-1} , берем математическое ожидание, делим на $\gamma(0)$.

Получаем:

$$\frac{E(X_t, X_{t-1})}{\gamma(0)} = \varphi_{21} \underbrace{\frac{E(X_{t-1}^2)}{\gamma(0)}}_{=1} + \varphi_{22} \frac{E(X_t, X_{t-1})}{\gamma(0)} + \frac{E(X_{t-1}, \varepsilon_t)}{\gamma(0)} \Rightarrow \rho_1 = \varphi_{21} + \varphi_{22}\rho_1.$$

$$\frac{E(X_t, X_{t-2})}{\gamma(0)} = \varphi_{21} \frac{E(X_{t-1}X_{t-2})}{\gamma(0)} + \varphi_{22} \frac{E(X_{t-1}^2)}{\gamma(0)} + \frac{E(X_{t-2}, \varepsilon_t)}{\gamma(0)} \Rightarrow \rho_2 = \varphi_{21}\rho_1 + \varphi_{22}.$$

В эти соотношение входят 2 неизвестных: φ_{21} и φ_{22} , а коэффициенты этой системы уравнений выражены через ρ_1 и ρ_2 , связь которых с коэффициентами исходного уравнения процесса нам уже известна. Другими словами, наш подход дает связь между коэффициентами автокорреляционной функции и частной автокорреляционной функции. В данном случае у нас всего 2 уравнения, можно решать их по-разному. Но поскольку нас интересует только один единственный коэффициент φ_{22} , наиболее экономно будет использовать правило Крамера.

$$\begin{cases} \varphi_{21} + \varphi_{22}\rho_1 &= \rho_1 \\ \varphi_{21}\rho_1 + \varphi_{22} &= \rho_2 \end{cases}$$

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \times \begin{bmatrix} \varphi_{21} \\ \varphi_{22} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \Rightarrow \varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{1 - \rho_1^2} \cdot \left(\frac{\Delta_2}{\Delta} \right)$$

Кстати, поскольку ρ_1 - это значение автокорреляционной функции, то оно меньше 1, то есть в знаменателе 0 не будет. Так как для процесса $AR(1)$ можно показать, что $\rho_k = \alpha^k \quad \forall k$, то

$$\varphi_{22} = \frac{\alpha^2 - \alpha^2}{1 - \rho^2} = 0 \quad , \text{ как и ожидалось.}$$

Для того чтобы посчитать φ_{33} , запишем теоретическую регрессию $X_t = \varphi_{31}X_{t-1} + \varphi_{32}X_{t-2} + \varphi_{33}X_{t-3} + \varepsilon_t$. Аналогично предыдущему получаем систему уравнений:

$$\begin{cases} \rho_1 &= \varphi_{31} \cdot 1 + \varphi_{32}\rho_1 + \varphi_{33}\rho_2 \\ \rho_2 &= \varphi_{31}\rho_1 + \varphi_{32} \cdot 1 + \varphi_{33}\rho_1 \\ \rho_3 &= \varphi_{31}\rho_2 + \varphi_{32}\rho_1 + \varphi_{33} \cdot 1 \end{cases}$$

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix}$$

Вновь нас интересует только коэффициент φ_{33} . Воспользуемся правилом Крамера, но, используя предыдущий опыт, считаем только определитель в числителе:

$$\Delta_3 = \begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix} = \begin{vmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha^2 \\ \alpha^2 & \alpha & \alpha^3 \end{vmatrix} = 0$$

(последний столбец пропорционален первому столбцу, причем коэффициент пропорциональности равен коэффициенту уравнения процесса) $\Rightarrow \varphi_{33} = 0$.

Теперь можно выразить определитель в числителе для произвольного k . Система линейных уравнений, связывающая значения автокорреляционной и частной автокорреляционной функций, фактически является системой уравнений Юла-Уокера для теоретической регрессии с неизвестными и подлежащими определению значениями частной автокорреляционной функции:

$$\begin{cases} \rho_1 &= \varphi_{k1} \cdot 1 + \varphi_{k2}\rho_1 + \varphi_{k3}\rho_2 + \dots + \varphi_{kk}\rho_{k-1} \\ \rho_2 &= \varphi_{k1}\rho_1 + \varphi_{k2} \cdot 1 + \varphi_{k3}\rho_1 + \dots + \varphi_{kk}\rho_{k-2} \\ \dots & \\ \rho_k &= \varphi_{k1}\rho_{k-1} + \varphi_{k2}\rho_{k-2} + \varphi_{k3}\rho_{k-3} + \dots + \varphi_{kk} \cdot 1 \end{cases}$$

Искомый определитель имеет вид:

$$\Delta_k = \begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_1 & \rho_2 \\ \vdots & \dots & \dots & 1 & \rho_{k-1} \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 & \rho_k \end{vmatrix}.$$

$$\Delta = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \vdots & \dots & \dots & \dots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{vmatrix}.$$

На главной диагонали все элементы кроме последнего будут равны 1. Связь между коэффициентами φ_{11} и $\rho_1, \dots, \rho_k$, то есть между автокорреляционной и частной автокорреляционной функциями, выражается следующими соотношениями:

$$\varphi_{kk} = \frac{\Delta_k}{\Delta}, \quad \text{где } \Delta - \text{определитель системы уравнений}$$

$$\Delta = \begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \dots & \dots & \rho_{k-2} \\ \vdots & \dots & \dots & \dots & \vdots \\ \rho_{k-1} & \dots & \dots & \dots & 1 \end{vmatrix}.$$

Таким образом, зная коэффициенты автокорреляционной функции, можно по уравнениям Юла-Уокера пересчитать коэффициенты частной автокорреляционной функции, и наоборот.

Если для $AR(1)$ считать четвертое, пятое и так далее значения частной автокорреляционной функции, то все они равны 0. Поэтому для процесса $AR(1)$ значения частной автокорреляционной функции, начиная со второго, равны 0 ($\varphi_{kk} = 0, \quad k > 1$). Следовательно, получен индикатор того, что исследуемый процесс точно является процессом $AR(1)$.

В общем случае процесса $AR(p)$: если $k > p$, то число столбцов в определителе Δ_k больше, чем порядок авторегрессии. В этом случае разностное уравнение $\rho_1, \dots, \rho_k$ показывает, что начиная с $k > p$, каждое ρ выражается одной и той же линейной комбинацией предыдущих значений. Как только число столбцов больше p , то каждый столбец с номером большим k является линейной комбинацией предыдущих столбцов. Причем коэффициенты этой линейной комбинации - это коэффициенты уравнения процесса $\alpha_1, \alpha_2, \dots, \alpha_p$. Таким образом, последний столбец всегда есть линейная комбинация p предыдущих столбцов, как только $k > p$. Поэтому определитель такой матрицы обязан обратиться в нуль.

Общий вывод: частная автокорреляционная функция авторегрессионного процесса $AR(p)$, равна 0 для $k > p$, и вообще говоря, не равна 0 при $k \leq p$.

В результате частная автокорреляционная функция для процесса AR играет точно такую же качественную роль, как автокорреляционная функция для процесса MA . Она обращается в нуль, как только $k > p$. Мы это доказали в общем случае и продемонстрировали на процессах первого и второго порядка.

Теперь делаем следующий шаг. Переходим к комбинации двух различных типов процессов $AR(p)$ и $MA(q)$. Рассмотрим общий процесс $ARMA(p, q)$ авторегрессионско-скользящего среднего:

$$x_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=0}^q \beta_j \varepsilon_{t-j}, \quad \beta_0 = 1, \text{ - в развернутом виде;}$$

$$\alpha_p(L)x_t = \beta_q(L)\varepsilon_t \quad \text{- в операторном виде.}$$

Чтобы выяснить, как ведет себя автокорреляционная и частная автокорреляционная функции процесса ARMA, начнем с простой ситуации, а именно с процесса $ARMA(1, 1)$.

В операторной записи: $(1 - \alpha L)x_t = (1 + \beta L)\varepsilon_t$, или в обычном виде:

$$x_t = \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$$

Условие стационарности процесса ARMA: $|\alpha| < 1$.

Для вычисления дисперсии процесса удобно использовать $MA(\infty)$ представление, так называемое представление линейного фильтра:

$$\begin{aligned} x_t &= \frac{1 + \beta L}{1 - \alpha L} \varepsilon_t = (1 + \beta L)(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots) \varepsilon_t = \\ &= [1 + (\alpha + \beta)L + \alpha(\alpha + \beta)L^2 + \alpha^2(\alpha + \beta)L^3 + \dots] \varepsilon_t. \end{aligned}$$

$$x_{t-1} = Lx_t$$

$$x_t = \alpha x_{t-1} + \beta \varepsilon_t + \beta \varepsilon_{t-1}$$

$$\begin{aligned} \frac{\gamma(1)}{\gamma(0)} &= \frac{\alpha\gamma(0) + \beta\sigma_\varepsilon^2}{\gamma(0)} = \alpha + \frac{\beta\sigma_\varepsilon^2}{\gamma(0)} = \alpha + \frac{\beta\sigma_\varepsilon^2(1 - \alpha^2)}{(1 + \beta^2 + 2\alpha\beta)\sigma_\varepsilon^2} = \\ &= \frac{\alpha + \alpha\beta^2 + 2\alpha^2\beta + \beta - \alpha^2\beta}{1 + \beta^2 + 2\alpha\beta} = \frac{\alpha + \beta + \alpha^2\beta + \beta + \alpha\beta^2}{1 + \beta^2 + 2\alpha\beta} = \\ &= \frac{(\alpha + \beta) + \alpha\beta(\alpha + \beta)}{1 + \beta^2 + 2\alpha\beta} = \frac{(\alpha + \beta) + (1 + \alpha\beta)}{1 + \beta^2 + 2\alpha\beta}. \end{aligned}$$

$$\begin{aligned} \text{Тогда } Var(x_t) &= \left[1 + (\alpha + \beta)^2 [1 + \alpha^2 + \alpha^4 + \dots] \right] \sigma_\varepsilon^2 = \left[1 + \frac{(\alpha + \beta)^2}{1 - \alpha^2} \right] \sigma_\varepsilon^2 = \\ &= \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma_\varepsilon^2 \Rightarrow \text{ автокорреляционная функция } \gamma(0) = \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma_\varepsilon^2. \end{aligned}$$

Чтобы посчитать первую автокорреляцию, умножим выражение $x_t = \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$ на x_{t-1} и возьмем математическое ожидание от обеих частей.

Получим $\gamma(1) = \alpha\gamma(0) + \beta Cov(\varepsilon_{t-1}, x_{t-1})$.

Умножив выражение $x_{t-1} = \alpha x_{t-2} + \varepsilon_{t-1} + \beta \varepsilon_{t-2}$ на ε_{t-1} и взяв математическое ожидание, получаем $Cov(\varepsilon_{t-1}, x_{t-1}) = \sigma_\varepsilon^2$. После подстановки получаем $\gamma(1) = \alpha\gamma(0) + \beta\sigma_\varepsilon^2$.

Окончательно:

$$\rho_1 = \frac{\gamma(1)}{\gamma(0)} = \frac{(\alpha+\beta)(1+\alpha\beta)}{1+\beta^2+2\alpha\beta}.$$

Последующие значения автокорреляционной функции получить проще. Если умножить x_t на x_{t-2} и взять математическое ожидание, получим: $\gamma(2) = \alpha\gamma(1)$. Для всех значений γ с индексом большим, чем на порядок MA -части, получаем, что из равенства $\gamma(k+1) = \alpha\gamma(k)$ следует равенство $\rho_{k+1} = \alpha\rho_k$. То есть мы получили то же самое соотношение, которое имели для "чистой" модели $AR(1)$. Мы называли это уравнениями Юла-Уокера для автокорреляционной функции. Другими словами, мы установили, что, начиная со второй, автокорреляции $ARMA(1, 1)$ ведут себя так же, как автокорреляции $AR(1)$, но первые автокорреляции этих процессов различаются.

Для $AR(1)$, автокорреляции имели вид $\rho_i = \alpha^i$. Для процесса $ARMA(1, 1)$ $\rho_1 \neq \alpha_1$, $\rho_1 = \frac{(\alpha+\beta)(1+\alpha\beta)}{1+\beta^2+2\alpha\beta}$, $\rho_i = \alpha\rho_{i-1}$, $i = 2, \dots$

Однако, начиная со второго номера, значения автокорреляционной функции все равно убывают экспоненциально.

Для процесса $ARMA(p, q)$ справедливо аналогичное утверждение, если, конечно, процесс стационарен. Первые p значений автокорреляционной функции определяются через коэффициенты AR и MA -частей, а потом значения автокорреляционной функции выражаются в виде суммы экспоненциально затухающих слагаемых.

Для процесса $ARMA(p, q)$, применяя тот же метод умножения уравнения процесса на значения x_{t-i} и последующего взятия математического ожидания, получим, что для $k > \max(p, q + 1)$, автокорреляции ρ_k определяются разностным уравнением, соответствующим AR -части. А все предыдущие значения ρ_k могут быть выражены через коэффициенты $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ как решения системы линейных уравнений, полученной способом аналогичным процессу $ARMA(1, 1)$. Следовательно, начиная с номера $\max(p, q + 1)$, автокорреляционная функция "затухает" том же смысле, что и для процесса AR . Таким же качественным поведением характеризуется и частная автокорреляционная функция процесса $ARMA(p, q)$. Интуитивно такое свойство понятно. Мы говорили, что для $MA(q)$ автокорреляционная функция для номеров, больших q , просто равна нулю. Поэтому влияние MA -части при $k > q$ как бы прекращается, и дальше "работает" только $AR(p)$. Наоборот, частная автокорреляционная функция для процесса $AR(p)$ для k , больших чем p , равна нулю. То есть $AR(p)$ перестает влиять на частную автокорреляционную функцию, и остается только влияние $MA(q)$.

Вывод: Если процесс относится к типу $ARMA(p, q)$, то, начиная с некоторого номера (причем этот номер важен, он нам говорит о величине p и q), и автокорреляционная, и частная автокорреляционная функции ведут себя как сумма затухающих экспонент, если ряд стационарен.

Пример 1.

Рассмотрим нестационарный ряд случайного блуждания. Его уравнение имеет вид: $x_t = x_{t-1} + \varepsilon_t$. Если взять первую разность, равную $\Delta x_t = x_t - x_{t-1} = (1 - L)x_t$, то уравнение сведется к следующему: $\Delta x_t = y_t = \varepsilon_t$. То есть в первых разностях ряд станет стационарным.

Такой подход приводит к стационарности не только случайное блуждание.

Пример 2.

Рассмотрим ряд вида: $x_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t$.

Его плотностью детерминированная часть (тренд) является параболической функцией времени. Очевидно, что этот ряд нестационарный. Математическое ожидание этого процесса зависит от времени.

Проверим: Приводится ли такой ряд взятием последовательных разностей к стационарному?

Если мы возьмем первую последовательную разность, то получим: $\Delta x_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t - \alpha - \beta(t-1) - \gamma(t-1)^2 \varepsilon_{t-1} = \beta + 2\gamma t - \gamma + (\varepsilon_1 - \varepsilon_{t-1})$.

Степень полинома, описывающего тренд, понизилась на единицу. Если провести взятие второй разности, то останется $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = 2\gamma + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2})$, то есть получим стационарный процесс.

Правда, видно, как в уравнение начинает "проникать" скользящее среднее. Полученный двукратным взятием разностей стационарный процесс является процессом $MA(2)$. Но, по крайней мере, взятием последовательных разностей исходный ряд с квадратичным трендом приводится к стационарному виду.

Бокс и Дженкинс на основании этого свойства предложили выделить класс нестационарных рядов, которые взятием последовательных разностей можно привести к стационарному виду, а именно к виду $ARMA$.

Если ряд после взятия d последовательных разностей приводится к стационарному, то этот ряд носит название $ARIMA(p, d, q)$.

$ARIMA$ - процесс авторегрессии - интегрированного скользящего среднего. При этом p - параметр AR -части, d - степень интеграции, и q - это параметр MA -части.

В операторном виде процесс $ARIMA(p, d, q)$ записывается как: $\alpha_p(L)\Delta^d x_t = \beta_q(L)\varepsilon_t$.

Или по-другому $\underbrace{\alpha_p(L)(1-L)^d}_{p+d} x_t = \beta_q(L)\varepsilon_t$.

Этот процесс нестационарный, потому что здесь не выполняется условие, что все корни характеристического уравнения по модулю меньше единицы.

Но, если обозначить $(1-L)^d x_t = y_t$, то y_t - это стационарный процесс.

Основная заслуга Бокса и Дженкинса в том, что они сделали этот подход лет 25 – 30 назад весьма популярным и ввели в практику программы для применения этого подхода в компьютерный пакет научных программ для системы $IBM - 360$, распространенной в 1970 – 1980гг. по всему миру.

Подход Бокса - Дженкинса

Бокс и Дженкинс предложили следующий подход, состоящий из нескольких этапов, к выбору модели типа $ARIMA$ по наблюдаемой реализации временного ряда.

1 этап "Идентификация модели"

1. Установить порядок интеграции d , то есть добиться стационарности ряда, взяв достаточное количество последовательных разностей. Другими словами, "остационарить" ряд. Используются все описанные ранее средства визуальный анализ коррелограммы визуальный анализ автокорреляционной и частной автокорреляционной функций, тест на единичный корень. $d \leq 2$ - обычно.
2. После этого мы получаем временной ряд y_t , которому нужно подобрать уже $ARMA(p, q)$. Исходя из поведения автокорреляционной и частной автокорреляционной функций, установить параметры p и q . Рекомендуется, чтобы $\gamma + \beta \leq 3$.

2 этап "Оценивание модели"

Оценивание коэффициентов $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ при условии, что мы уже знаем p и q .

3 Этап "Диагностика модели"

Стандартная для эконометрического подхода процедура. По остаткам осуществляется тестирование или диагностика построенной модели.

4 Этап "Использование модели"

Использование модели, в основном, для прогнозирования будущих значений временного ряда.

Бокс и Дженкинс применили этот подход ко многим временным рядам, которые были известны в то время, как к финансовым, так и к макроэкономическим.

Бокс и Дженкинс смогли построить модели типа *ARIMA* для всех исследуемых рядов и установили, в частности, что практически все экономические процессы описываются моделями с относительно небольшими величинами параметров p и q , а параметр d обычно не превышает 2. К тому же оказалось, что точность прогнозирования по моделям *ARIMA* оказалась выше, чем давали в то время экономические модели.

Если исследуемый ряд нестационарный, то его автокорреляционная функция не будет убывать.

Если ряд стационарен, то, начиная с какого-то номера, теоретические автокорреляции будут убывать. Поэтому можно рассчитать их оценки - выборочные автокорреляции, и посмотреть, убывают они или нет. Если ряд окажется стационарным, перейти к определению параметров p и q . Если нет, то надо построить ряд первых разностей и проверить на стационарность его.

Рассмотрим, например, процесс $x_t = \alpha x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim WN(0, \sigma^2)$.

При $\alpha > 1$ взятие первой разности не поможет сделать ряд стационарным. Это нестационарный ряд взрывного типа, оценки его "автокорреляционной" функции растут с увеличением сдвига во времени.

Если же $\alpha = 1$, то ряд представляет собой случайное блуждание и после взятия первой разности он станет стационарным.

Переходим к оцениванию по выборке статистических характеристик исследуемого процесса в предположении его эргодичности. В качестве оценки математического ожидания применяется статистика $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$, то есть обычное среднее по выборке.

В качестве оценки дисперсии процесса обычно принимается следующая величина: $S^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2$. Обратите внимание, делитель не $(T-1)$, как привычно для обработки независимых наблюдений, а T .

Для оценки коэффициента теоретической автокорреляции используем $\hat{\rho}_k = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{TS^2}$.

Это означает, что для расчета выборочных ковариаций используется одинаковый делитель - T , что, в случае независимых наблюдений, дает смещенные оценки соответствующих теоретических ковариаций. Если бы мы требовали несмещенности оценок, то у нас бы были разные делители при различных k . Кроме того, при увеличении k уменьшается объем выборки, пригодный для расчета оценки соответствующего коэффициента корреляции. На практике, по рекомендации, идущей еще от Бокса и Дженкинса, не рассчитываются оценки $\hat{\rho}_k$ для $k > \frac{T}{4}$.

Одним из следствий выбора именно таких оценок значений автокорреляционной функции является гарантированная положительная определенность выборочной корреляционной матрицы. Ранее мы отмечали наличие такого свойства у теоретической автокорреляционной функции. При одинаковом делителе у оценок автокорреляционной функции при различных k положительная определенность выборочной корреляционной матрицы гарантирована. Поэтому мы предпочитаем пусть смещенную оценку, но гарантирующую это свойство, тем более, что у нас обычно T - относительно большое. Во-вторых, при дополнительном предположении о нормальности распределения значений процесса именно эта оценка совпадает с оценкой метода максимального правдоподобия. Оценка $\hat{\theta}$ параметра θ несмещенная, если $E(\hat{\theta}) = \theta$.

I этап "Идентификация модели"

Некоторые выводы, полученные ранее:

1. Если при k большем некоторого q выборочные автокорреляции $\hat{\rho}_k$ становятся близкими к нулю, то подходящей моделью может быть $MA(q)$.
2. Если при k большем некоторого p частные выборочные автокорреляции $\hat{\phi}_{kk}$ становятся близкими к нулю, то подходящей моделью может быть $AR(p)$.
3. Если и $\hat{\rho}_k$ и $\hat{\phi}_{kk}$ стремятся к нулю плавно, затухают по экспоненте, то подходящей моделью является смешанная модель $ARMA(p, q)$.

Важным является также принцип экономичности: предпочтительны модели с небольшими значениями p и q .

Для уточнения того, что понимать под словами " ϕ_{kk} близки к нулю" или " ϕ_{kk} близки к нулю" могут быть использованы следующие результаты.

Бокс и Дженкинс (1976) показали следующее.

Если x_t - это процесс $MA(q)$ и процесс белого шума ε_t распределен по нормальному закону, то для $k > q$ и при больших T выборочная корреляция r_k асимптотически распределена по нормальному закону со следующими параметрами:

$$\begin{aligned}\bar{r}_k &= E(r_k) = 0, \\ S_r^2 &= Var(r_k) \approx \frac{1}{T} \text{ при } k = 1, \\ Var(r_k) &\approx \frac{1}{T}(1 + 2 \sum_{j=1}^{k-1} r_j^2) \text{ при } k > 1.\end{aligned}$$

(каждой реализации случайного процесса при любом k соответствует свое значение ρ_k).

Этот результат также используется для оценки дисперсии путем замены ρ_k на ϕ_{kk} .

Гипотеза $H_0 : \rho_k = 0$ для $k > q$ означает, что процесс x_t - есть процесс $MA(q)$.

Если x_t - это процесс $AR(p)$, то аналогичное утверждение справедливо для частных выборочных автокорреляций ϕ_{kk} , то есть $H_0 : \phi_{kk} = 0$ при $k > p$ ϕ_{kk} асимптотически распределена по нормальному закону со следующими параметрами:

$$\begin{aligned}E(\phi_{kk}) &= 0, \\ Var(\phi_{kk}) &\approx \frac{1}{T}, \\ (\text{при } k > p \sqrt{T}\phi_{kk} &\sim N(0, 1)).\end{aligned}$$

II этап "Оценивание коэффициентов моделей типа ARMA"

Рассмотрим модель $AR(p) : \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \varepsilon_t$. Для нее можно применить обычный метод наименьших квадратов (МНК).

Поскольку регрессоры относятся к предыдущим моментам времени ($t - j \neq t$), а ε_t - белый шум, то корреляция регрессоров x_{t-j} со случайным возмущением ε_j отсутствует.

Метод наименьших квадратов (МНК) дает 1) смещенные оценки $E\hat{\theta} \neq \theta$; 2) состоятельные оценки $\lim_{n \rightarrow \infty} \hat{\theta} \rightarrow \theta$, несмотря на то, что присутствует стохастический регрессор.

Нужно проверить по остаткам, действительно ли наши предположения о том, что ε_t - белый шум, выполнены.

Если дополнительно белый шум является гауссовым ($\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$), то значения x_t распределены нормально, а оценки коэффициентов $\hat{\alpha}_1, \dots, \hat{\alpha}_p$ произведенные МНК, состоятельны и асимптотически нормальны.

Если данные имеют ненулевое выборочное среднее $\bar{x}$, то можно либо вычесть это среднее из данных и строить регрессию без свободного члена, т.е. рассматривать $x_t - \bar{x}$, либо просто строить регрессию со свободным членом θ .

Другими словами, для оценки математического ожидания процесса $AR(p)$ можно использовать две статистики: уже упомянутую в прошлой лекции $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$

и $\hat{\mu} = \frac{\hat{\theta}}{1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p}$, в которой использованы оценки МНК. Для гауссова процесса ε_t обе оценки состоятельны и асимптотически нормальны. Более того, обе они асимптотически независимы от оценок параметров модели, полученных МНК.

Для моделей скользящего среднего $MA(q)$ ($x_t = \sum_{j=1}^q \beta_j \varepsilon_{t-j}$) невозможно аналитически выразить остаточную сумму квадратов через значения реализации x_t и параметры модели, а следовательно, применить МНК. Для оценки параметров существуют два подхода, по-разному реализованные сегодня в специализированных компьютерных программах.

Применение методов максимального правдоподобия (ММП)

В предположении нормальности ошибки ε_t выражаем ковариационную матрицу ошибок (ее элементы - значения автокорреляционной функции) через параметры $\beta_1, \beta_2, \dots, \beta_q$ обратной модели $MA(q)$. Функция правдоподобия для нормального распределенного вектора $\vec{x} = [x_1, x_2, \dots, x_T]'$ имеет вид

$$L(\vec{\beta}, \sigma_\varepsilon^2 \sum_x |\vec{x}) = \frac{1}{(\sqrt{2\pi\sigma_\varepsilon^2})^T} \left(\sqrt{\det \sum_x} \right)^{-1} \exp \left(- \frac{(\vec{x} - \bar{x}) \sum_x^{-1} (\vec{x} - \bar{x})}{2\sigma_\varepsilon^2} \right)$$

$$\varepsilon \mathcal{N}(0, \sigma_\varepsilon^2 \sum_x)$$

Здесь через $\sum_x$ обозначена ковариационная матрица процесса $\vec{x}$. Элементы этой матрицы выражаются, согласно свойству обратимости, через параметры модели $\beta_1, \beta_2, \dots, \beta_q$, поэтому процедура численной оптимизации (максимизации функции правдоподобия) позволяет найти оценки метода максимального правдоподобия, которые будут обладать обычными свойствами состоятельности и асимптотической нормальности. Кроме того, оценки параметров модели $MA(q)$ и оценка дисперсии случайного возмущения асимптотически независимы.

Применение процедуры нелинейной оптимизации поиска на сетке (grid-search procedure).

Этот подход, позволяющий облегчить вычислительные затраты, применили Бокс и Дженкинс.

В общем случае модели $MA(q)$ 1) Фиксируем начальные значения $\beta_1^0, \dots, \beta_q^0$ и определяем остатки через наблюдения: $e_1 = x_1 - \bar{x}$, $e_2 = x_2 - \bar{x} - \beta_1^0 e_1$ И так далее. Процедура получения остатков достаточно проста, все остатки с отрицательными и нулевым индексами в используемых выражениях заменяем нулями. 2) Затем находим $\min_{\vec{\beta}} \sum e_i^2$. 3) Численным поиском находим оценки параметров.

Рассмотрим идею этого подхода на примере процесса $MA(2)$: $x_t = \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$.

1. Найдем сначала оценку математического ожидания процесса $\bar{x} = E x_t$ в предположении о его стационарности и эргодичности.
2. Затем выберем некоторые значения (β_1, β_2) , например $(0, 5; 0, 5)$.
3. Исходя из уравнения процесса $(x_k = \bar{x} + \beta_1 e_{k-1} + \beta_2 e_{k-2} + e_k)$, x_1 выражается через предыдущие ошибки, но они не известны. Поэтому будем считать, что

$$x_1 = \bar{x} + e_1,$$

$$x_2 = \bar{x} + \beta_1 e_1 + e_2,$$

$$x_3 = \bar{x} + \beta_1 e_2 + \beta_2 e_1 + e_3,$$

$$x_4 = \bar{x} + \beta_1 e_3 + \beta_2 e_2 + e_4$$

Это дает нам $e_1 = x_1 - \bar{x}$. Затем полагаем

$$e_2 = x_2 - \bar{x} - \beta_1 e_1,$$

$$e_3 = x_3 - \bar{x} - \beta_1 e_2 - \beta_2 e_1$$

$$e_4 = x_4 - \bar{x} - \beta_1 e_3 - \beta_2 e_2$$

и так далее.

4. Теперь можно вычислить сумму квадратов отклонений $\sum e^2 = S(\beta_1, \beta_2)$. Поскольку она посчитана для выбранных значений параметров (β_1, β_2) , мы можем рассматривать ее как функцию этих переменных.
5. Далее какой-либо численной процедурой, например поиском на сетке, можно перебирать комбинации β_1, β_2 или численно искать минимум функции $\min_{\beta_1, \beta_2} \sum e_t^2$.

Коэффициенты (β_1^*, β_2^*) , которые обеспечивают минимум выражения $\sum e_t^2$ и будут оценками коэффициентов модели $MA(2)$.

Замечание: Была доказана асимптотическая эквивалентность этой процедуры оцениванию методом максимального правдоподобия (при $T \rightarrow \infty$).

Для оценки параметров модели $ARMA(p, q)$ может применяться комбинация метода наименьших квадратов с поиском на сетке.

Рассмотрим модель $ARMA(2, 2)$.

Пусть уравнение модели имеет вид

$$(1 - \alpha_1 L - \alpha_2 L^2)x_t = (1 + \beta_1 L + \beta_2 L^2)\varepsilon_t$$

Его можно переписать в виде

$$x_t = \frac{(1 + \beta_1 L + \beta_2 L^2) \varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$$

Введем вспомогательный случайный процесс $x_t = \frac{\varepsilon_t}{1 - \alpha_1 L - \alpha_2 L^2}$. Тогда процессы x_t и z_t связаны соотношением $x_t = (1 + \beta_1 L + \beta_2 L^2) z_t$, которое напоминает уравнение $MA(2)$, только вместо ε_t стоит z_t .

Определим "наблюдения" z_t через наблюдения x_t , сконструировав z_t так же, как ранее остатки модели МА, т.е. значения z_t , которые еще не определены (с нулевыми и отрицательными индексами), полагаем равными нулю. Тогда:

$$z_1 = x_1 \quad (x_1 = z_1 + \beta_1 z_0 (= 0) + \beta_2 z_{-1} (= 0))$$

$$z_2 = x_2 - \beta_1 z_1 \quad (x_2 = z_2 + \beta_1 z_1 + \beta_2 z_0 (= 0))$$

$$z_3 = x_3 - \beta_1 z_2 - \beta_2 z_1$$

...

Далее из определения процесса z_t следует, что значения z_t связан с остатками e_t исходной модели следующими соотношениями: $z_t - \alpha_1 z_{t-1} - \alpha_2 z_{t-2} = e_t$. Относительно процесса z_t модель стала $AR(2)$.

Зная "реализацию" z_t для выбранных значений коэффициентов (β_1^0, β_2^0) , можно оценить коэффициенты α_1, α_2 с помощью МНК.

В результате находим оценки коэффициентов $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)$, но получены они по-разному. Оценки $(\tilde{\beta}_1, \tilde{\beta}_2)$ заданы как начальные значения, т.е. $(\tilde{\beta}_1, \tilde{\beta}_2) = (\beta_1^0, \beta_2^0)$. А потом, исходя из них, построены оптимальные оценки $(\tilde{\alpha}_1, \tilde{\alpha}_2)$.

Применяя численные методы оптимизации (например, поиск на сетке) оцениваем значения параметров, обеспечивающие $\min_{\alpha_1, \alpha_2, \beta_1, \beta_2} \sum e_i^2$.

Если ε_t - белый гауссовский шум, то **для оценивания модели $ARMA(p, q)$** можно также применять **метод максимального правдоподобия**.

Общая схема его применения следующая.

1. Выражаем значения автокорреляционной функции процесса через параметры модели $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$
2. Строим ковариационную матрицу порядка
3. Записываем функцию правдоподобия для имеющейся выборки $x_1, \dots, x_T$
4. Решаем (как правило, численно) систему уравнений правдоподобия относительно оценок коэффициентов $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_q$

Реальные алгоритмы метода максимального правдоподобия, реализованные в специализированных компьютерных пакетах, отличаются от этой схемы, но принципиально она осуществима и отражает логику получения оценок ММП.

После того, как параметры модели оценены, остается оценить качество полученной модели.

III Этап "Диагностика модели типа ARMA"

Исходной информацией для диагностики служат остатки модели. Поскольку предполагалось, что случайное возмущение ε_t является белым шумом, поэтому, прежде всего, надо проверять некоррелированность остатков.

Таким образом, проверке подлежат, в первую очередь, 2 момента.

Первый момент - это **качество модели**. Для проверки качества модели нужен индикатор типа критерия множественной детерминации R^2 .

Второй момент - это **некоррелированность остатков**. Если остатки коррелированы, то оценивать $AR(p)$ -часть методом наименьших квадратов нельзя, так как получим несостоятельные оценки.

Проверка качества модели

Как правило, при построении моделей временных рядов критерии качества подгонки моделей применяются для сравнения моделей между собой. Поскольку оценки коэффициентов проводятся путем оптимизации, фактически речь идет о выборе порядка модели, т.е. о сравнении моделей с различным числом параметров.

Абсолютные критерии, типа стандартного коэффициента множественной детерминации R^2 , не применяются.

Очевидно, что модели с большими p и q дают лучшие (меньшие) значения суммы квадратов остатков, чем модели с меньшими p и q . Но это противоречит принципу экономности.

Наиболее распространенными в настоящее время является предложенный Акаике в 1974 г. **критерий AIC** (Akaike information criterion) и предложенный Шварцем в 1978 г. **критерий BIC** (Bayesian information criterion). Оба эти критерия построены примерно одинаковым способом.

Информационный критерий AIC для модели $ARMA(p, q)$ выглядит следующим образом:

$$AIC(p, q) = \ln \hat{\sigma}^2 + 2 \frac{p + q}{T}$$

где T - число наблюдений. **Байесовский информационный критерий BIC** имеет несколько другой вид:

$$BIC(p, q) = \ln \hat{\sigma}^2 + \frac{p + q}{T} \ln T$$

Здесь $\hat{\sigma}^2 = \frac{\sum e_t^2}{T - p - q}$ (другой вариант $\hat{\sigma}^2 = \frac{\sum \varepsilon_t^2}{T}$).

Структура этих критериев следующая: логарифмы остаточной суммы квадратов плюс штраф за уменьшение числа степеней свободы.

Нужно так подобрать значения параметров p и q , чтобы получить минимальное значение каждого из критериев: $\min_{p, q} AIC(p, q)$, $\min_{p, q} BIC(p, q)$, в то время как по критерию R^2 мы отдавали предпочтение модели с большим его значением.

Замечание: В отличие от коэффициентов детерминации ничего нельзя сказать ни о диапазоне изменения, ни о знаке критериев Акаике и Шварца.

Критерий Акаике базируется на обобщении ММП. Приведенное выражение подразумевает, что случайное возмущение является гауссовым. Шибата показал, что для процессов авторегрессии критерий AIC переоценивает порядок модели, и оценка порядка модели на основании этого критерия несостоятельна.

Критерий Шварца ВИС основан на байесовском подходе и имеет более фундаментальное теоретическое обоснование. Показано, что оценка порядка модели по этому критерию является состоятельной.

Тем не менее, на практике чаще используется информационный критерий АИС. Существует масса работ, в которых сравнивается применение этих критериев по отношению к разным моделям, но окончательного вывода пока не сделано. На практике разные критерии могут привести к выбору различных моделей. Есть некоторый накопленный опыт по поводу того, к каким типам моделей приводит один критерий и к каким приводит другой. Если вы публикуете какие-то исследование, считается вполне нормальным просто указать, какой критерий вы применяете без специального обоснования вашего выбора.

Проверка автокорреляции остатков

В этом вопросе существует расхождение между тем, как следовало бы поступать с теоретической точки зрения и тем, что делается практически.

В 1970 г. Бокс и Пирс предложили **статистический критерий для проверки автокорреляции временного ряда**, который сегодня принято называть **Q-статистикой**.

Тест Бокса-Пирса проверяет гипотезу о совместном равенстве нулю всех автокорреляций временного ряда до порядка m включительно, т.е. гипотезу $H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$ против альтернативной гипотезы $H_1 : \sum_{i=1}^m \rho_i^2 > 0$.

Бокс и Пирс показали, что при увеличении длины выборки, т.е. при больших T , статистика $Q(m) = T \sum_{k=1}^m r_k^2$ (r_k - выборочные автокорреляции) имеет асимптотическое распределение χ^2 с m степенями свободы.

Этот тест было предложено применять для проверки наличия автокорреляции остатков типа $ARMA(p, q)$. В этом случае число степеней свободы уменьшается на $(p + q)$, т.е. статистика $Q(m)$ имеет асимптотическое распределение χ^2 с $m - p - q$ степенями свободы, а если при этом ε_t является процессом нормального белого шума, то $Q(m)$ имеет распределение χ^2 с $m - p - q$ степенями свободы.

Работа Бокса и Пирса была опубликована в 1970г. В том же году Дарбин показал, что статистика Дарбина-Уотсона не применима для проверки автокорреляции остатков при наличии в качестве регрессора объясняемой переменной с лагом. Одновременно Дарбин предложил так называемую h -статистику и альтернативную процедуру Дарбина для проверки наличия автокорреляции в таких моделях.

Позднее было замечено, что в моделях с лаговой объясняемой переменной в качестве регрессора статистика Бокса-Пирса не применима по тем же причинам, что и статистика Дарбина-Уотсона. Однако, поскольку в моделях временных рядов проверка наличия автокорреляции критически важна, статистика Бокса-Пирса нашла широкое применение и до сих пор входит в состав большинства специализированных эконометрических пакетов.

Так как статистика Бокса-Пирса имеет малую мощность, Бокс и Льюнг в 1978 г. предложили использовать для тех же целей **улучшенную Q-статистику**:

$$Q = (T + 2)T \sum_{i=1}^m (T - i)^{-1} r_j^2$$

По сравнению со статистикой Бокса-Пирса различным слагаемым приданы разные Веса. Авторы доказали, что эта статистика имеет то же асимптотическое распределение χ_m^2 , но лучше им аппроксимируется при конечном числе наблюдений.

Статистика Бокса-Льюнга теоретически не применима для тестирования автокорреляции остатков в моделях ARMA по тем же причинам, что и статистики Дарбина-Уотсона и Бокса-Пирса. Тем не менее, статистика Бокса-Льюнга входит во все специализированные пакеты, множество исследователей ею пользуются, хотя она теоретически не состоятельна.

Q -статистику Бокса-Пирса и улучшенную Q -статистику Бокса-Льюнга часто называют **портманто-статистикой** (portmanteau-statistics).

Пример. При помощи 100 нормальных случайных чисел $\{\varepsilon_t\}$ было построено 100 значений $\{y_t\}$:

$$y_t = -0,7y_{t-1} + \varepsilon_t - 0,7\varepsilon_{t-1}$$

y_0 и ε_0 , были приняты равными нулю.

Происхождение данных считается неизвестным, и сравниваются три следующие модели:

Модель 1 ($AR(1)$): $y_t = \alpha_1 y_{t-1} + \varepsilon_t$

Модель 2 ($ARMA(1, 1)$): $y_t = \alpha_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$

Модель 3 ($AR(2)$): $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$

Проверка автокорреляции остатков. Результаты расчетов таковы:

$$Q_{0,95}(8) = 15,51 \quad Q_{0,95}(24) = 36,42$$

Модель 1: $\alpha_1 = -0,835$

$$Q(8) = 26,19(0,000) \quad AIC = 507,3$$

$$Q(24) = 41,10(0,001) \quad BIC = 509,9$$

Модель 2: $\alpha_1 = -0,679 \quad \beta_1 = -0,676$

$$Q(8) = 3,86(0,695) \quad AIC = 481,4$$

$$Q(24) = 14,23(0,892) \quad BIC = 486,6$$

Модель 3: $\alpha_1 = -1,16 \quad \alpha_2 = -0,378$

$$Q(8) = 11,44(0,057) \quad AIC = 492,5$$

$$Q(24) = 22,59(0,424) \quad BIC = 497,7$$

Модель 1 должна быть отброшена по значениям статистик $Q(8)$ и $Q(24)$. Предпочтение должно быть отдано модели 2 перед моделью 3 по значениям обоих информационных критериев, а также потому, что в модели 3 значение статистики $Q(8)$ указывает на наличие определенной корреляции между остатками.

Q -статистика (Бокса-Пирса) и улучшенная Q -статистика (Бокса-Льюнга) не применимы для временных рядов с лаговой переменной в авторегрессионной части, но практически они используются для проверки автокорреляционных остатков.

Для проверки наличия автокорреляции в моделях ARMA лучше воспользоваться более мощным и универсальным способом, а именно **методом множителей Лагранжа** (Lagrange multiplier - LM), применительно к проверке автокорреляции остатков его еще называют тестом Бройша-Годффри (Breusch-Godfrey).

Он входит «триаду» классических асимптотических тестов: отношения правдоподобий, Вальда, множителей Лагранжа, и применим для широкого класса задач проверки ограничений на коэффициенты модели.

Пусть рассматривается модель множественной регрессии

$$y_t = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon_t$$

где $x_1, \dots, x_k$ разные регрессоры, в том числе, возможно, и лаговые значения как объясняющих, так и объясняемой переменных.

Проверим предположение, что ε_t подчиняется авторегрессионной схеме порядка p , т.е. задается уравнением $\varepsilon_t = \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_p \varepsilon_{t-p} + u_t$, где u_t - белый шум.

В терминах коэффициентов модели основная и альтернативная гипотезы принимают вид $H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$, $H_1 : \sum_{i=1}^p \gamma_i^2 > 0$

Метод множителей Лагранжа для проверки этой гипотезы заключается в следующем. Методом МНК строится обычная регрессия вида $y_t = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon_t$. Обозначим ее остатки через e_t .

1. Строится регрессия либо той же объясняемой переменной y_t , либо остатков e_t на старые регрессоры и остатки с лагом до p включительно (т.е. в качестве дополнительных объясняющих переменных используем $e_{t-1}, e_{t-2}, \dots, e_{t-p}$).

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k + \hat{\gamma}_1 e_{t-1} + \hat{\gamma}_2 e_{t-2} + \dots + \hat{\gamma}_p e_{t-p}$$

2. Проверяется гипотеза о том, что группа дополнительных объясняющих переменных является излишней. Если в качестве объясняемой переменной используются остатки, то статистика TR^2 имеет асимптотическое (при увеличении числа наблюдений) распределение χ^2 с p степенями свободы (T - число наблюдений, R^2 - коэффициент множественной детерминации регрессии, построенной в пункте 1). Если $TR^2 \chi_p^2$ при $T \rightarrow \infty$, то принимаем гипотезу.

Для проверки гипотезы о равенстве нулю группы переменных обычно мы привыкли использовать F-статистику, проверяющую, фактически, статистическую значимость уменьшения остаточной суммы квадратов от включения дополнительных переменных. Разумеется, F-статистика применима и в этом случае. Но, она применима только при нормальном распределении случайного члена ε_t . Применение же теста множителей Лагранжа не требует нормальности распределения, но «работает» только асимптотически.

Современные эконометрические пакеты, в частности Eviews, сообщают пользователю при применении LM-теста значения и критические вероятности (p-values) для обеих статистик (F и TR^2), так что можно выбирать: использовать ли F-отношение, верное для конечных выборок, но в предположении нормальности, либо не требовать нормальности и использовать статистику TR^2 , верную лишь асимптотически.

Для проверки нормальности существует множество тестов. Рассмотрим тест Харке-Бера (Jarque-Bera).

Этот тест вычисляет выборочные значения для коэффициентов асимметрии $S = \frac{1}{T\sigma^3} \sum (e_t - \bar{x})^3$ и эксцесса $K = \frac{1}{T\sigma^4} \sum (e_t - \bar{x})^4$, где $\bar{x}$ - выборочное среднее, а σ - выборочное среднеквадратичное отклонение остатков модели.

При условии нормальности остатков, статистика Харке-Бера $\frac{T-p-q-1}{6} \left[S^2 + \frac{1}{4}(K-3)^2 \right]$ имеет χ^2 распределение с двумя степенями свободы.

IV этап "Прогнозирование с помощью моделей ARMA"

Прогнозирование будущих значений экономической величины является одним из основных способов применения моделей временных рядов. Использование моделей типа

ARMA обладает некоторыми особенностями по сравнению с прогнозированием по модели множественной регрессии.

При прогнозировании по правильно специфицированной модели существует 2 источника ошибок прогноза:

- неопределенность будущих значений случайной величины ε
- отсутствие точных значений коэффициентов модели $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$ (у нас есть только их оценки $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_q$, оцененные по имеющейся выборке).

При прогнозировании по модели множественной регрессии мы оцениваем значение зависимой переменной y_t при заданных значениях независимых переменных - регрессоров $x_{t1}, x_{t2}, \dots, x_{tk}$. Это приводит к тому, что характеристики прогноза, как случайной величины, являются, по сути, условными характеристиками при условии имеющейся выборки независимых переменных.

Иная ситуация в моделях типа ARIMA. Здесь значение переменной x_t прогнозируется для некоторого будущего момента времени $t + \tau$, при этом лаговые значения этой переменной $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-p}$, служащие регрессорами модели, можно рассматривать фиксированными на выборочных значениях, или случайными. Первая возможность, когда $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-p}$ фиксированы, приводит к **условному прогнозу**, как и для модели множественной регрессии, а вторая, когда $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-p}$ случайные, - к **безусловному прогнозу**. Таким образом, при прогнозировании по модели типа ARIMA можно рассматривать как условный, так и безусловный прогнозы. Из курса теории вероятностей известно, что условная дисперсия случайной величины не превышает ее безусловную дисперсию, поэтому точность условного прогноза всегда выше.

Условный прогноз:

прогноз на 1 шаг:

$$\tilde{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1 \varepsilon_T + \dots + \beta_q \varepsilon_{T-q+1} | x_1, \dots, x_T\} = \theta + \beta_1 l_T + \dots + \beta_q l_{T-q+1}$$

прогноз на 2 шага:

$$\tilde{x}_{T+2} = E\{\theta + \varepsilon_{T+2} + \beta_1 \varepsilon_{T+1} + \dots + \beta_q \varepsilon_{T-q+2} | x_1, \dots, x_T\} = \theta + \beta_2 l_T + \dots + \beta_q l_{T-q+2}$$

прогноз на q шагов:

$$\tilde{x}_{T+q} = E\{\theta + \varepsilon_{T+q} + \beta_1 \varepsilon_{T+q-1} + \dots + \beta_q \varepsilon_T | x_1, \dots, x_T\} = \theta + \beta_q l_T$$

$\tilde{x}_{T+q+1} = E\{\theta + \varepsilon_{T+q+1} + \beta_1 \varepsilon_{T+q} + \dots + \beta_q \varepsilon_{T+1} | x_1, \dots, x_T\} = \theta$ - условный прогноз совпадает с безусловным

Тест Харке — Бера

Материал из Википедии — свободной энциклопедии

Тест Харке—Бера (англ. *Jarque-Bera test*) — это статистический тест, проверяющий ошибки наблюдений на нормальность посредством сверки их третьего момента (асимметрия) и четвёртого момента (эксцесс) с моментами нормального распределения, у которого $S = 0, K = 3$.

В тесте Харке—Бера проверяется нулевая гипотеза $\mathcal{H}_0: S = 0, K = 3$ против гипотезы $\mathcal{H}_1: S \neq 0, K \neq 3$, где S — коэффициент асимметрии (Skewness), K — коэффициент эксцесса (Kurtosis)

Формулировка

Тест выглядит следующим образом:

$$JB = n \left(\frac{S^2}{6} + \frac{(K-3)^2}{24} \right), \text{ где } S = \frac{\sum e_i^3}{n \hat{\sigma}_{ML}^3}, K = \frac{\sum e_i^4}{n \hat{\sigma}_{ML}^4}, e_i \text{ — остатки модели, } n \text{ — ко-}$$

личество наблюдений, $\hat{\sigma}_{ML}^2 = \frac{\sum e_i^2}{n}$, ML — обозначение метода максимального правдоподобия (Maximal Likelihood). Данная статистика имеет распределение хи-квадрат с двумя степенями свободы (χ_2^2), поскольку коэффициенты S и K асимптотически нормальны, следовательно, их квадраты при нормировке дадут две случайные величины, распределённые как χ_1^2 . Чем ближе распределение ошибок к нормальному, тем меньше статистика Харке—Бера отличается от нуля. При достаточно большом значении статистики p -value будет мало, и тогда будет основание отвергнуть нулевую гипотезу (статистика попала в «хвост» распределения).

Свойства теста

Тест Харке—Бера является *асимптотическим* тестом, то есть применим к большим выборкам. Если ошибки распределены нормально, то в соответствии с теоремой Гаусса—Маркова оценки метода наименьших квадратов будут лучшими (иметь наименьшую дисперсию в классе линейных несмещённых оценок), и коэффициенты регрессии будут также распределены асимптотически нормально.

Литература

Damodar N. Gujarati. Basic Econometrics. — 4. — The McGraw-Hill Companies, 2004. — С. 1002. — ISBN 978-0071123433.

Источник - "https://ru.wikipedia.org/wiki/Тест_Харке_—_Бера"

При прогнозировании по модели ARIMA от имеющейся выборки зависят как оценки коэффициентов модели $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_q$, так и значения регрессоров $x_t, x_{t-1}, x_{t-2}, \dots, x_{t-p}$, поэтому сложно аналитически выразить условную дисперсию ошибки прогноза через имеющиеся значения временного ряда. Общепринято ограничиваться не очень реалистичным предположением о том, что коэффициенты модели известны точно. Разумеется, это предположение уменьшает дисперсию ошибки прогноза, чем увеличивает кажущуюся точность как условного, так и безусловного прогнозов.

Показано, что если мы хотим достичь минимума среднеквадратической ошибки (MSE), т.е. не требовать несмещённости ($E\{\hat{\theta}\} \neq \theta$), то надо взять условное математическое ожидание: $E\{x_{T+\tau}|x_1, \dots, x_T\}$. Такое условное математическое ожидание будет гарантировать получение MSE, или иногда ее обозначают MMSE (minimum mean square error).

Рассмотрим **модель МА(q)**: $x_t = \theta + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$.

Полагаем, что:

1. коэффициенты модели точно известны;
2. имеются значения x_t для $t \in [1; T]$.

Очевидно, что безусловным точечным прогнозом для любого момента времени будет математическое ожидание процесса, т.е. θ ($\hat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1 \varepsilon_T + \dots + \beta_q \varepsilon_{T-q+1}\} = \theta$). Условным прогнозом для момента времени $T + 1$ будет условное математическое ожидание:

$$\hat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1 \varepsilon_T + \dots + \beta_q \varepsilon_{T-q+1} | x_1, \dots, x_T\}.$$

Среди случайных величин ε , которые стоят в левой части, есть такие, которые связаны с имеющимися наблюдениями. Ведь наблюдение складывается из "модельного" значения и ошибки, поэтому условные математические ожидания всех слагаемых, кроме ε_{T+1} , не равны нулю: $E\{\varepsilon_{T+1}|x_1, \dots, x_T\} = 0$, $E\{\varepsilon_T|x_1, \dots, x_T\} \neq 0$, $E\{\varepsilon_{T-q+1}|x_1, \dots, x_T\} \neq$

0 ($E\{y|x\} = E\{y\}$, если x и y независимы). Рассмотрим, например, $E\{\varepsilon_T|x_1, \dots, x_T\}$. Это математическое ожидание - остаток между наблюдением и прогнозом по модели, т.е. $E\{\varepsilon_T|x_1, \dots, x_T\} = e_T = x_T - \hat{x}_T$. Поэтому условные математические ожидания от всех предыдущих значений случайной составляющей надо заменить соответствующими остатками. Следовательно, условные математические ожидания для прошлых значений - остатки, для будущих значений - нули.

Точно так же конструируется прогноз не на 1, а на 2 и вообще на τ шагов вперед. Все последующие ε заменяются нулями, а предыдущие - заменяются реально наблюдаемыми остатками. Таким образом, для модели $MA(q)$ прогноз зависит от того, какие ошибки были на предыдущих шагах. Начиная с шага $(q + 1)$ условный прогноз представляет собой просто математическое ожидание θ , т.е. условный прогноз совпадает с безусловным: $\forall \tau : \tau \geq q + 1$

$\tilde{x}_{T+\tau} = E\{x_{T+\tau}\} = \theta$, следовательно q - кратная память об ошибках.

Рассмотрим условную дисперсию ошибки прогноза на 1 шаг:

$$Var\{x_{T+1} - \hat{x}_{T+1}|x_1, \dots, x_T\} = E\{(\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + \dots + \beta_q\varepsilon_{T-q+1} - \theta - \beta_1e_T - \dots - \beta_qe_{T-q+1}|x_1, \dots, x_T)^2\} = \sigma_\varepsilon^2$$

Остатки определяются из уравнения $x_i = \hat{x}_i + e_i$.

Аналогично дисперсия прогноза на 2 шага равна $Var\{x_{T+2} - \hat{x}_{T+2}|x_1, \dots, x_T\} = (1 + \beta_1^2)\sigma_\varepsilon^2$, а дисперсия прогноза на τ шагов составит $(1 + \beta_1^2 + \dots + \beta_\tau^2)\sigma_\varepsilon^2$ при $\tau < q$. При $\tau \geq q$ дисперсия ошибки условного прогноза становится равной дисперсии ошибки безусловного прогноза, т.е. просто дисперсии случайного процесса x_t .

Теперь рассмотрим **модель стационарного процесса $AR(p)$** :

$$x_t = \theta + \alpha_1x_{t-1} + \dots + \alpha_px_{t-p} + \varepsilon_t.$$

Для прогноза на 1 шаг вперед можно записать:

$$Var\{x_{T+1} - \tilde{x}_{T+1}|x_1, \dots, x_T\} = E\{a_{T+1}^2|x_1, \dots, x_T\} - E^2\{a_{T+1}|x_1, \dots, x_T\}$$

$$a = x_{T+1} - \tilde{x}_{T+1}$$

$$E\{a_{T+1}^2|x_1, \dots, x_T\} = E\{(\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + \dots + \beta_q\varepsilon_{T-q+1} - \theta - \beta_1e_T - \dots - \beta_qe_{T-q+1}|x_1, \dots, x_T)^2\} = E\{(\varepsilon_{T+1}|x_1, \dots, x_T)^2\} = \sigma_\varepsilon^2$$

$$E^2\{a_{T+1}|x_1, \dots, x_T\} = E^2\{\varepsilon_{T+1}|x_1, \dots, x_T\} = 0$$

прогноз на 2 шага:

$$Var\{x_{T+2} - \tilde{x}_{T+2}|x_1, \dots, x_T\} =$$

$$E\{(x_{T+2} - \tilde{x}_{T+2}|x_1, \dots, x_T)^2\} - E^2\{x_{T+2} - \tilde{x}_{T+2}|x_1, \dots, x_T\} = \{\text{второе слагаемое равно нулю}\}$$

$$= E\{(\theta + \varepsilon_{T+2} + \beta_1\varepsilon_{T+1} + \dots + \beta_q\varepsilon_{T-q+2} - \theta - \beta_1e_T - \dots - \beta_qe_{T-q+2}|x_1, \dots, x_T)^2\} =$$

$$E\{(\varepsilon_{T+2} + \beta_1\varepsilon_{T+1}|x_1, \dots, x_T)^2\} = (1 + \beta_1^2)\sigma_\varepsilon^2$$

прогноз на τ шагов:

$$Var\{x_{T+\tau} - \tilde{x}_{T+\tau}|x_1, \dots, x_T\} = (1 + \beta_1^2 + \dots + \beta_\tau^2)\sigma_\varepsilon^2, \text{ если } \tau < q$$

$$Var\{x_{T+\tau} - \tilde{x}_{T+\tau}|x_1, \dots, x_T\} = Var\{x_{T+\tau} - \tilde{x}_{T+\tau}\}, \text{ если } \tau \geq q.$$

$$Var\{x_{T+1} - \hat{x}_{T+1}|x_1, \dots, x_T\} = Var\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + \dots + \beta_q\varepsilon_{T-q+1} - \theta - \hat{b}_1e\}$$

$$x_{T+1} - \hat{x}_{T+1} = x_{T+1} - (x_T - e_T)$$

$$\hat{x}_{T+1} = E\{\theta + \varepsilon_{T+1} + \beta_1\varepsilon_T + \dots + \beta_q\varepsilon_{T-q+1}|x_1, \dots, x_T\}$$

$$Var\{x|z\} := Cov(x, x|z) = E\{(x - E(x|z))^2|z\} = E\{(x_{T+1} - \hat{x}_{T+1}) - E(x_{T+1} - \hat{x}_{T+1}|x_1, \dots, x_T)\}$$

$$E\{x_{T+2}|x_1, \dots, x_T\} = E\{\theta + a_1x_{T+1} + \dots + a_px_{T-p} + \varepsilon_{T+2}|x_1, \dots, x_T\} = \theta + a_1\hat{x}_{T+1} + \dots + a_px_{T-p+1}$$

$$\lim_{\tau \rightarrow \infty} E\{x_{T+\tau}|x_1, \dots, x_T\} = \frac{\theta}{\theta - \alpha_1 - \dots - \alpha_p}$$

$$x_t = a_1x_{t-1} + a_2x_{t-2} + \varepsilon_t$$

Условную дисперсию ошибки прогноза можно рассчитать аналогично случаю модели скользящего среднего, но выкладки становятся весьма громоздкими, даже для моделей малого порядка.

Рассмотрим, например, модель $AR(2)$ без свободного члена. Тогда $\hat{x}_{T+1} = \alpha_1x_T + \alpha_2x_{T-1}$, а $x_{T+1} = \alpha_1x_T + \alpha_2x_{T-1} + \varepsilon_{T+1}$. Очевидно, что дисперсия ошибки прогноза на 1 шаг равна σ_ε^2 . Для прогноза на 2 шага соответственно получаем:

$$\tilde{x}_{T+2} = \alpha_1\tilde{x}_{T+1} + \alpha_2x_T = \alpha_1(\alpha_1x_T + \alpha_2x_{T-1}) + \alpha_2x_T,$$

$$x_{T+2} = \alpha_1x_{T+1} + \alpha_2x_T + \varepsilon_{T+2} = \alpha_1(\alpha_1x_T + \alpha_2x_{T-1} + \varepsilon_{T+1}) + \alpha_2x_T + \varepsilon_{T+2},$$

Дисперсия ошибки прогноза на 2 шага равна: $(1 + \alpha_1^2)\sigma_\varepsilon^2$.

Для прогноза на 3 шага получим:

$$\tilde{x}_{T+3} = \alpha_1\tilde{x}_{T+2} + \alpha_2\tilde{x}_{T+1} = \alpha_1(\alpha_1(\alpha_1x_T + \alpha_2x_{T-1}) + \alpha_2x_T) + \alpha_2(\alpha_1x_T + \alpha_2x_{T-1}),$$

$$x_{T+3} = \alpha_1x_{T+2} + \alpha_2x_{T+1} + \varepsilon_{T+3} = \alpha_1(\alpha_1(\alpha_1x_T + \alpha_2x_{T-1} + \varepsilon_{T+1}) + \alpha_2x_T + \varepsilon_{T+2}) + \alpha_2(\alpha_1x_T + \alpha_2x_{T-1} + \varepsilon_{T+1}) + \varepsilon_{T+3}.$$

Дисперсия ошибки прогноза на 3 шага равна $(1 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1^2\alpha_2 + \alpha_1^4)\sigma_\varepsilon^2$.

Очевидно, что дисперсия ошибки увеличивается от шага к шагу. Значительно более простые выражения для дисперсии ошибки прогноза получаются, если перейти от $AR(p)$ представления модели к эквивалентному МА представлению $x_t = \theta + \varepsilon_t + \psi_1\varepsilon_{t-1} + \dots + \psi_q\varepsilon_{t-q} + \dots$, хотя и с бесконечным числом слагаемых. Тогда дисперсия

ошибки прогноза на τ шагов выражается формулой $\sigma_\varepsilon^2 \sum_{i=0}^{\tau-1} \psi_i^2 (\psi_0 = 1)$. (*)

Для общей **модели ARMA(p, q)** нужно просто объединить полученные результаты. Если мы хотим получить прогноз с минимальной среднеквадратической ошибкой, то:

1) рассчитываем прогнозные значения $x_{T+\tau}$ по нашей модели, подставляя туда для времени $[1, T]$ - наблюдаемые значения $x_1, \dots, x_T$ и рассчитанные значения остатков $e_1, \dots, e_T$, а для всех последующих моментов времени $\tau = T+1, T+2, \dots$ - заменяем остатки нулями ($e_{T+1} = 0, e_{T+2} = 0, \dots$), а для значений $x_{T+1}, x_{T+2}, \dots$ подставляем их прогнозные значения $\tilde{x}_{T+1}, \tilde{x}_{T+2}, \dots$

2) Для получения дисперсии ошибки прогноза переходим от ARMA к МА представлению и пользуемся формулой (*).

$$(*)Var\{x_{T+1} - \tilde{x}_{T+1}|x_1, \dots, x_T\} = Var\{a_1x_T + \alpha_2x_{T-1} + \varepsilon_{T+1} - a_1x_T - \alpha_2x_{T-1}|x_1, \dots, x_T\} = \sigma_\varepsilon^2$$