Подбор и диагностика модели АРСС(n, m)

Родионов Максим Владимирович, группа 522

Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — доц., канд. физ.-мат. наук **Товстик Т.М.**
Рецензент — доц., канд. физ.-мат. наук **Сизова А.Ф.**

Санкт-Петербург
2007г.

Основные понятия и определения

- Стохастическое уравнение АРСС(n, m):

$$x_t + \sum_{j=1}^{n} q_j x_{t-j} = \sum_{k=0}^{m} p_k \xi_{t-k},$$

$$\mathbf{E}\xi_j = 0, \ \ \mathbf{E}\xi_k \xi_j = \delta_{kj}.$$

- Случайный процесс $x(t)$ называется **стационарным в широком смысле (слабо стационарным)**, если выполнены следующие условия:
  1. $\mathbf{E}\,|x(t)|^2 < +\infty \quad \forall\, t \in \mathcal{T}$;
  2. $m(t) \equiv m$ (математическое ожидание не зависит от времени);
  3. автоковариация зависит лишь от разности аргументов: $R(t, s) = R(t - s)$.
- Характеристическое уравнение: $q(t) = 1 + q_1 t + \ldots + q_n t^n$.

Основные понятия и определения

- Случайный процесс x_t называется **обратимым**, если существует эквивалентное представление процесса x_t в виде процесса авторегрессии бесконечного порядка АР(∞)

$$x_t = \sum_{j=1}^{\infty} d_j x_{t-j} + \xi_t,$$

$$d(z) = 1 - \sum_{j=1}^{\infty} d_j z^j = \frac{q(z)}{p(z)},$$

$$q(z) = \sum_{j=0}^{n} q_j z^j, \qquad p(z) = \sum_{j=0}^{m} p_j z^j.$$

- Характеристическое уравнение: $p(t) = p_0 + p_1 t + \ldots + p_m t^m$.

Постановка задачи

Мы предполагаем, что некоторый наблюдаемый временной ряд порождается моделью АРСС(n, m), при этом возникает проблема подбора конкретной модели из этого класса и оценивания ее коэффициентов.

- ***Задача*:**
  построить алгоритм, который решает данную проблему в три этапа:
  - **Идентификация модели,**
  - **Оценивание модели,**
  - **Диагностика модели.**

Идентификация модели

- На этапе идентификации модели производится выбор некоторой частной модели из всего класса АРСС(n, m), т.е. выбор значений n и m.
- Отправная точка — различие поведения частных автокорреляционных (ρ^{part}) и автокорреляционных (ρ) функций рядов, соответствующих различным моделям АРСС.
- ρ_k^{part} определяется как решение относительно q_k системы первых k уравнений Юла-Уолкера:

$$q_1\rho_{s-1} + q_2\rho_{s-2} + \ldots + q_k\rho_{s-k} = \rho_s, \quad s = 1, 2, \ldots, k,$$

$$\rho_k^{part} = q_k.$$

- Выборочные автокорреляции:

$$\hat{\rho}_k = \frac{\frac{1}{N-k}\sum_{t=1}^{N-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\frac{1}{N}\sum_{t=1}^{N}(x_t - \overline{x})^2} = \frac{\hat{R}_k}{\hat{R}_0}$$

Идентификация модели

По правилу Крамера:

$$\rho_0^{part} = 1,$$

$$\rho_1^{part} = \rho_1,$$

$$\vdots$$

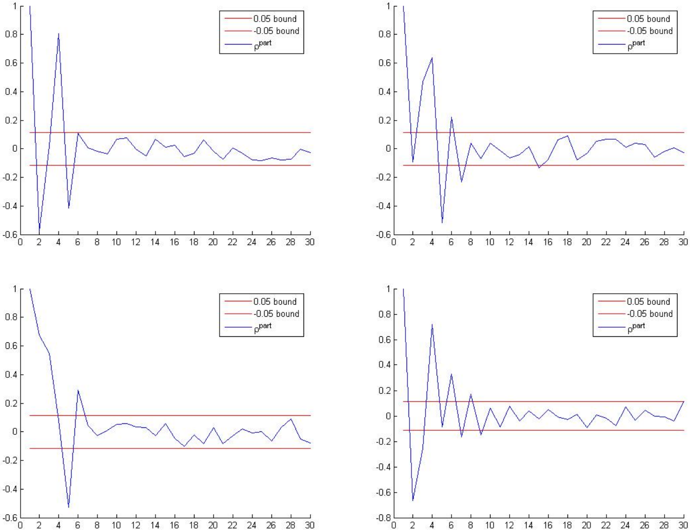
$$\rho_k^{part} = \frac{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_2 \\ \rho_2 & \rho_1 & \dots & \rho_{k-4} & \rho_3 \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \rho_2 & \rho_1 & \dots & \rho_{k-4} & \rho_{k-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{vmatrix}}.$$

Идентификация модели

- Гипотезу $H_0 : x_t \sim$ АР (n) будем отвергать при
$$|\rho_k^{part}| > \frac{2}{\sqrt{N}}, \quad \forall\, k > n.$$
- Гипотезу $H_0 : x_t \sim$СС(m) будем отвергать при
$$|\rho_k| > \frac{2}{\sqrt{N}}, \quad \forall\, k > m.$$
- Определение параметра n АРСС(n, m):
ρ_{part} — осциллирующее или прямое убывание, начинающееся с лага n.
- Определение параметра m АРСС(n, m):
ρ — осциллирующее или прямое убывание, начинающееся с лага m.

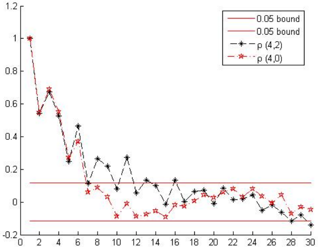
Идентификация модели

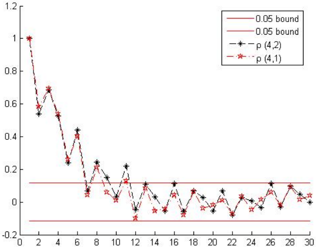
Графики частных автокорреляций АРСС$(4, m)$, $m = 0, \ldots, 3$.

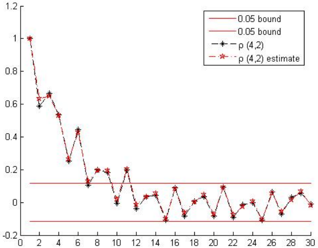


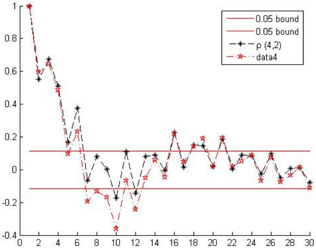
Идентификация модели

Графики автокорреляций АРСС$(4, m)$, $m = 0, \ldots, 3$.









Оценивание параметров процесса

Оценка коэффициентов процесса АРСС по заданным наблюдениям $x_1, x_2, \ldots, x_N$, образующим временной ряд.
Сначала оценим коэффициенты $q_1, \ldots, q_n$ одним из следующих методов:

1. Решаем систему уравнений Юла — Уолкера:

$$q_1\rho_{s-1} + q_2\rho_{s-2} + \ldots + q_n\rho_{s-n} = -\rho_s, \quad s = m+1, \ldots, m+n.$$

2. Решаем систему уравнений:

$$\sum_{j=0}^{n} q_j \sum_{k=m+1}^{v} \rho_{k-j}\rho_{k-t} = 0, \quad q_0 = 1, \quad t = 1, \ldots, n.$$

Оценки параметров $q_1, \ldots, q_n$ получим при замене автоковариаций ρ_j их оценками $\hat{\rho}_j$.

Оценивание параметров процесса

Теперь требуется оценить коэффициенты $p_0, \ldots, p_m$. Для этого рассмотрим вспомогательный процесс y_j, определяемый следующим образом:

$$y_t = \sum_{k=0}^{n} \hat{q}_k x_{t-k}, \quad \hat{q}_0 = 1, \quad t = 1, \ldots, N.$$

Нахождение оценок для $p_0, \ldots, p_m$ сводится к решению системы нелинейных уравнений

$$\begin{array}{rcl} p_0^2 + p_1^2 + \ldots + p_m^2 & = & G_0, \\ p_0 p_1 + p_1 p_2 + \ldots + p_{m-1} p_m & = & G_1, \\ \ldots & \ldots & \ldots, \\ p_0 p_m & = & G_m. \end{array}$$

где G_j — это корреляции процесса y_t. В качестве G_j нужно использовать их оценки

$$\hat{G}_j = \sum_{k=0}^{n} \sum_{l=0}^{n} \hat{q}_k \hat{q}_l \hat{\rho}_{|j+k-l|}.$$

Диагностика оцененной модели

Этот этап позволяет понять насколько хорошо подобранная модель соответствует данным наблюдениям.

- стационарность при $\hat{q}_i, \ i = 0, \ldots, n$,
- обратимость при $\hat{p}_j, \ j = 0, \ldots, m$.

Рассмотрим величину σ, которую будем вычислять по формуле

$$\sigma = \frac{\sqrt{\sum_{k=0}^{\nu} |\rho_k - \hat{\rho}_k|^2}}{\nu + 1},$$

$$\nu \geq n + m.$$

Соответственно, чем меньше σ, тем лучше подобранный ряд описывает исследуемую модель.

Пример

Промоделируем модель АРСС$(4,2)$ и попытаемся подобрать под нее модель АРСС$(4,m)$, $m = 0,\ldots,3$:

Таблица: Истинные значения параметров АРСС$(4,2)$.

АРСС	q_1	q_2	q_3	q_4	p_0	p_1	p_2	p_3
$(4,2)$	0.23	-0.7	-0.53	0.4	1.23	0.84	0.41	–

Таблица: Оценка параметров АРСС$(4,m)$, $m = 0,\ldots,3$.

АРСС	q_1	q_2	q_3	q_4	p_0	p_1	p_2	p_3	σ
$(4,0)$	-0.24	-0.77	-0.22	0.52	1.35	–	–	–	0.32
$(4,1)$	0.12	-0.89	-0.42	0.54	1.18	0.77	–	–	0.22
$(4,2)$	0.32	-0.62	-0.58	0.33	1.21	1.1	0.45	–	0.19
$(4,3)$	0.063	-0.82	-0.43	0.6	0.16	1.68	1.25	-0.26	–

Заключение

- На этапе идентификации был произведен выбор некоторой частной модели из всего класса АРСС(n, m), т.е. выбор n и m. Хотелось бы сразу отметить, что 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 этапе оценивания произведена оценка коэффициентов модели с использованием эффективных статистических методов. Разработка данного этапа показала значимую роль размера выборки N и снижение точности методов с ростом n и m.
- На третьем этапе проверена адекватность выбранной модели имеющимся данным. Неадекватности, обнаруженные в процессе диагностики модели, указывают на необходимую корректировку модели, после чего производится новый цикл подбора, и т.д. до тех пор, пока не будет получена удовлетворительная модель.