Эконометрика. Лекция 8. Модели временных рядов

12 декабря 2014 г.

Временные ряды:

- Многомерные

(тут табличка)

- Одномерные

Одномерный временной ряд

Временной ряд — последовательность случайных величин

$$y_1, y_2, y_3, \ldots$$

Без предположений невозможно прогнозировать

1, 2, 3, 4, 5, ?

(потом появляется правильный ответ: 42)

Базовое предположение — стационарность

Временной ряд называется стационарным, если:

- $E(y_1) = E(y_2) = E(y_3) = \ldots$
- $Var(y_1) = Var(y_2) = Var(y_3) = \ldots = \gamma_0$
- $Cov(y_1, y_2) = Cov(y_2, y_3) = Cov(y_3, y_4) = \ldots = \gamma_1$
- $Cov(y_1, y_3) = Cov(y_2, y_4) = Cov(y_3, y_5) = \ldots = \gamma_2$
- $\ldots$

Предпосылки коротко:

Временной ряд называется стационарным, если:

- $E(y_t) = const$
- $Cov(y_t, y_{t-k}) = \gamma_k$

Автоковариационная функция

$\gamma_k = Cov(y_t, y_{t-k})$ — (авто)-ковариационная функция процесса

Самый простой пример — белый шум

Ряд ε_t — белый шум, если:

- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma^2$
- $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$

Пример белого шума

$\varepsilon_t \sim N(0, 4)$ и независимы

(график)

Конвенция об обозначениях

На эту лекцию ε_t всегда обозначает белый шум!

Примеры нестационарного процесса

- Процесс с детерминистическим трендом
- Случайное блуждание

Процесс с детерминистическим трендом

- $y_t = 5 + 6t + \varepsilon_t$.

(тут график)

Здесь: $E(y_t) = 5 + 6t \neq const$

Случайное блуждание

- $\begin{cases} y_0 = 0 \\ y_t = y_{t-1} + 2 + \varepsilon_t \end{cases}$

(тут график)

Здесь: $Var(y_t) = t\sigma^2$

Процесс скользящего среднего

Процесс представимый в виде

$$y_t = \mu + \varepsilon_t + a_1 \varepsilon_{t-1} + \dots a_q \varepsilon_{t-q}$$

Обозначение процесса скользящего среднего

$y_t \sim MA(q)$, Moving Average

Чудо доска. Пример MA процесса

$$y_t = 5 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

Найдите $E(y_t)$, $Var(y_t)$, $Cov(y_t, y_{t-k})$

Запись с помощью оператора лага

L — оператор лага:

- $Ly_t = y_{t-1}$
- $L^2 y_t = y_{t-2}$
- ...

Пример записи с помощью оператора лага

$MA(2)$:

$$y_t = 2 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

$$y_t = 2 + (1 + 3L - 2L^2)\varepsilon_t$$

Интерпретация:

Коэффициенты плохо интерпретируемы

У стационарного процесса:

$\rho_k = Corr(y_t, y_{t-k})$ — (авто)-корреляционная функция процесса

Интерпретация

Если y_t — стационарный процесс и $y_t \sim N(\mu_y, \sigma_y^2)$, то:

ρ_k — на сколько в среднем изменится y_t при росте y_{t-k} на единицу

Чудо доска. Автокорреляционная функция MA процесса

$$y_t = 5 + \varepsilon_t + 3\varepsilon_{t-1} - 2\varepsilon_{t-2}$$

Частная автокорреляционная функция-идея

ϕ_k — частная автокорреляционная функция

ρ_k — автокорреляционная функция

(тут картинка со стрелочками)

Частная автокорреляционная функция-интерпретация

Если y_t — стационарный процесс и $y_t \sim N(\mu_y, \sigma_y^2)$, то:

ϕ_k — на сколько в среднем изменится y_t при росте y_{t-k} на единицу при фиксированных y_{t-1}, y_{t-2}, $\ldots$, y_{t-k+1}

Частная автокорреляционная функция-определение

$$\phi_k = Cor(y_t - P(y_t), y_{t-k} - P(y_{t-k}))$$

где $P(y_t)$ — проекция случайной величины y_t на линейную оболочку величин y_{t_1}, y_{t-2}, …, y_{t-k+1}.

Частная автокорреляция алгоритм подсчёта

$$\phi_k = det(A_k^*)/det(A_k)$$

- A_k — ковариационная матрица вектора $(y_1, y_2, \ldots, y_k)$
- A_k^* — матрица A_k, в которой последний столбец заменили на $(\gamma_1, \gamma_2, \ldots, \gamma_k)$.

Чудо доска. Частная Автокорреляционная функция МА процесса

$$y_t = 5 + \varepsilon_t + 3\varepsilon_{t-1}$$

Процесс авторегрессии

- Стационарный процесс вида

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \varepsilon_t$$

Обозначение процесса авторегрессии

$y_t \sim AR(p)$, AutoRegression

Чудо-доска. Частная и обычная автокорреляционные функции для AR процесса

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t \; \varepsilon_t \sim N(0,1)$$

Найдите ρ_k, ϕ_k

Альтернативная форма записи

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t$$

Или

$$(y_t - 4) = 0.5(y_{t-1} - 4) + \varepsilon_t$$

Важное предупреждение

Из одного уравнения $y_t = 2 + 0.5y_{t-1} + \varepsilon_t$ не следует автоматически стационарность (!)

Чудо-доска. Пример множества решений

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t \ \varepsilon_t \sim N(0,1)$$

- $y_0 = 0$, $y_1 \sim N(2,1)$, $y_2 \sim N(3,1.25)$, ...
- $y_0 \sim N(3,4/3)$, $y_1 \sim N(3,4/3)$, $y_2 \sim N(3,4/3)$, ...

Подразумеваем стационарное решение

Пишем:

$$y_t = 2 + 0.5y_{t-1} + \varepsilon_t \; \varepsilon_t \sim N(0,1)$$

Подразумеваем:

- $y_0 \sim N(3, 4/3)$, $y_1 \sim N(3, 4/3)$, $y_2 \sim N(3, 4/3)$, ...

AR процесс можно записать с помощью лага

$$y_t = 2 + 0.5y_{t-1} - 0.06y_{t-2} + \varepsilon_t$$

$$(1 - 0.5L + 0.06L^2)y_t = 2 + \varepsilon_t$$

Характеристический многочлен

$$(1 - 0.5L + 0.06L^2)y_t = 2 + \varepsilon_t$$

$$f(L)y_t = 2 + \varepsilon_t$$

$f(L)$ — характеристический многочлен

Когда y есть стационарное решение?

$$f(L)y_t = +\varepsilon_t$$

Если корни характеристического уравнения AR процесса по модулю меньше единицы, то существует единственное стационарное решение, в котором y_t выражается через прошлые шумы, то есть через ε_t, ε_{t-1}, ε_{t-2}, ...

Чудо-доска

Два примера AR(2): стационарный и нет

Прогнозирование

Прогноз на h шагов вперед: $E(y_{t+h}|y_t, y_{t-1}, y_{t-2}, \ldots)$

Часто кратко обозначают: $\hat{y}_{t+h}$

Чудо-доска

$$y_t = 2 + 0.5y_{t-1} - 0.06y_{t-2} + \varepsilon_t \; \varepsilon_t \sim N(0;4)$$

$y_{100} = 4$, $y_{99} = 3$.

постройте точечный и интервальный прогноз на 1 и 2 шага вперед

Модель авторегрессии и скользящего среднего

- Стационарный процесс вида

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \varepsilon_t + a_1 \varepsilon_{t-1} + \ldots + a_q \varepsilon_{t-q}$$

где сумма $p + q$ минимально возможна

Обозначение

- $y_t \sim ARMA(p, q)$

Сумма $p + q$ минимально возможная

- $y_t = \varepsilon_t$
- $y_t - y_{t-1} = \varepsilon_t - \varepsilon_{t-1}$

Вывод: $y_t \sim ARMA(0,0)$

ARMA — это наше всё!

Теорема. Любой стационарный процесс можно представить в виде $AR(\infty)$

Вывод. С помощью $ARMA(p, q)$ можно компактно и очень точно описать любой стационарный процесс

Итого про ARMA(p,q)

- коэффициенты не интерпретируемы
- используются для прогнозирования

Оценивание коэффициентов

Есть T наблюдений: $y_1, y_2, y_3, \ldots, y_T$

Чаще всего используется метод максимального правдоподобия

Подробности метода максимального правдоподобия

- Предполагается независимость и нормальность $\varepsilon_t \sim N(0;\sigma^2)$
- Стационарность y_t

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + \varepsilon_t + a_1 \varepsilon_{t-1} + \ldots + a_q \varepsilon_{t-q}$$

Результат метода максимального правдоподобия

На выходе получаем оценки

$$\hat{\theta} = (\hat{a}_1, \ldots, \hat{a}_q, \hat{b}_1, \ldots, \hat{b}_q, \hat{\sigma}^2)$$

И оценку их ковариационной матрицы $\widehat{Var}(\hat{\theta})$

Проверка гипотез и доверительные интервалы

$$\frac{\hat{a}_j - a_j}{se(\hat{a}_j)} \to N(0;1)$$

Выборочная автокорреляционная функция

ACF — autocorrelation function

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Выборочная частная автокорреляционная функция

PACF — partial autocorrelation function

Получим $\hat{\phi}_k$ из оценки регрессии

$$\hat{y}_t = * + * \cdot y_{t-1} + * \cdot y_{t_2} + \ldots + * \cdot y_{t-k+1} + \phi_k y_{t-k} + u_t$$

Примечания к расчету автокорреляционной функции

- Для оценки каждого $\hat{\phi}_k$ строится отдельная регрессия
- Из каждой регрессии нужен только последний коэффициент

Алгоритм на практике

1. Графики ряда, ACF, PACF
2. Если ряд нестационарный, то преобразуем
3. Выбираем p и q
4. Оцениваем $ARMA(p, q)$
5. Прогнозируем

Основное преобразование

Взятие разности: переход от y_t к Δy_t

Обозначение

- $y_t \sim ARIMA(p, 1, q)$ равносильно $\Delta y_t \sim ARMA(p, q)$
- $y_t \sim ARIMA(p, 0, q)$ равносильно $y_t \sim ARMA(p, q)$

Выбор p и q по графикам

График выборочной корреляционной функции есть даже у нестационарного процесса!

Белый шум

(график)

Случайное блуждание (нестационарный процесс!)

(график)

Процесс с трендом (нестационарный процесс!)

(график)

AR(1) и AR(2)

два графика

MA(1) и MA(2)

два графика

ARMA(1,1)

график

Мораль

- временные ряды: стационарные и нет
- для стационарных — модель ARMA