

# Учебник по временным рядам: начало

Винни-Пух

30 января 2023 г.

## Содержание

|                                |    |
|--------------------------------|----|
| Белый шум, стационарность и МА | 1  |
| Частная корреляция             | 4  |
| Строго про сходимости          | 5  |
| Оператор лага                  | 5  |
| Разностные уравнения и ARMA    | 6  |
| Стационарный бутстрэп          | 7  |
| ETS                            | 8  |
| Сглаживание ряда               | 8  |
| Скользящее среднее             | 8  |
| LOESS                          | 8  |
| Выделение сезонности           | 9  |
| STL                            | 9  |
| MSTL                           | 10 |
| Байесовские модели             | 10 |
| DLT                            | 10 |
| Тесты на прогнозную силу       | 14 |
| Тест Диболда-Мариано           | 14 |
| RC и SPA тесты                 | 15 |
| Решения                        | 15 |
| Источники мудрости             | 15 |

## Белый шум, стационарность и МА

Из курса математического анализа мы знаем разницу между рядами и последовательностями. В последовательности числа записаны одно за другим, скажем, через запятую,

$$5, 8, -3, 2, 4, 5, \dots$$

А ряд — это бесконечная сумма чисел, например,

$$0.9 + 0.09 + 0.009 + 0.0009 + \dots$$

Настала пора дать первое определение и раскрыть заговор рептилоидов!

**Определение 1.** Временной ряд — это последовательность случайных величин.

Индекс временного ряда может быть любым, но чаще всего мы работаем с тремя случаями. Бесконечный в обе стороны индекс,

$$\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots$$

бесконечный в одну сторону,

$$y_1, y_2, y_3, y_4, \dots$$

либо конечный,

$$y_1, y_2, y_3, y_4, \dots, y_T.$$

Чтобы отличать весь временной ряд от одной конкретной случайной величины, мы будем использовать обозначения:

$y_t$  — одна конкретная случайная величина;

$(y_t) = y_1, y_2, y_3, y_4, \dots$  — вся последовательность случайных величин.

Если контекст требует, то можно проявить больше аккуратности и указать возможные значения индекса, например,  $(y_t)_{t=1}^{\infty}$ .

Начнём с самого простого временного ряда — белого шума.

**Определение 2.** Ряд  $(u_t)$  называется белым шумом (white noise), если он удовлетворяет трём свойствам:

- Нулевое математическое ожидание,  $E(u_t) = 0$  для любого  $t$ .
- Постоянная дисперсия,  $\text{Var}(u_t) = \sigma_u^2$  для любого  $t$ .
- Нулевая ковариация,  $\text{Var}(u_t, u_s) = 0$  для любых  $t \neq s$ .

Заметим, что случайные величины в белом шуме вполне могут быть зависимы. Например,

...

**Определение 3.** Ряд  $(y_t)$  называется слабо стационарным (weakly stationary), или просто стационарным, если он удовлетворяет трём свойствам:

- Постоянное математическое ожидание,  $E(y_t) = \mu$  для любого  $t$ .
- Постоянная дисперсия,  $\text{Var}(y_t) = \gamma_0$  для любого  $t$ .

Да, да, всё верно, временной ряд — это не ряд, ноль — чётное число, единица — не простое, булённые кубики — не кубики, московские диаметры — не диаметры, а Деда Мороза не существует.

или занудства

в) Ковариация двух величин зависит только от их удалённости по времени друг от друга,  $\text{Var}(y_t, y_s) = \gamma_{t-s}$  для любых  $t$  и  $s$ .

Из третьего условия на ковариацию  $\text{Var}(y_t, y_s) = \gamma_{t-s}$  следует постоянство дисперсии, достаточно подставить  $t = s$  и увидеть, что  $\text{Var}(y_t, y_t) = \text{Var}(y_t) = \gamma_0$ . Мы выписали второе свойство отдельно от третьего, чтобы лучше его выделить.

**Определение 4.** Ряд  $(y_t)$  называется процессом скользящего среднего порядка  $q$  (moving average of order  $q$ ) относительно белого шума  $(u_t)$ , если он представим в виде:

$$y_t = \mu + u_t + \alpha_1 u_{t-1} + \dots + \alpha_q u_{t-q},$$

где последний коэффициент  $\alpha_q \neq 0$ .

Обозначаем такие процессы мы так:  $y_t \sim MA(q)$ .

Процесс скользящего среднего — это статистическая модель. Название скользящего среднего имеет одна из простых процедур сглаживания ряда.

**Определение 5.** Ряд  $(y_t)$  называется процессом скользящего среднего бесконечного порядка (moving average of infinite order) относительно белого шума  $(u_t)$ , если он представим в виде:

$$y_t = \mu + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots,$$

где ...

Обозначаем такие процессы мы так:  $y_t \sim MA(\infty)$ .

В определении мы не требуем, чтобы бесконечное количество коэффициентов  $\alpha_i$  были отличны от нуля. Вполне возможно, что после некоторого номера  $q$  все последующие  $\alpha_i = 0$ , поэтому и белый шум, и  $MA(q)$  процессы являются частными случаями  $MA(\infty)$  процесса.

Заметим, что один и тот же процесс  $(y_t)$  может быть представлен по-разному относительно разных белых шумов.

Приведём несколько примеров.

**Пример 1.** Посмотрим на  $MA(\infty)$  со специально подобранными коэффициентами

$$y_t = u_t + \dots$$

Попробуем посчитать ковариации

$$\text{Var}(y_t, y_{t+1}) =$$

$$\text{Var}(y_t, y_{t+2}) =$$

Считая дальнейшие ковариации по аналогии, обнаруживаем, что  $(y_t)$  — белый шум! Другими словами, один белый шум может являться нетривиальной линейной комбинацией значений другого белого шума.

**Пример 2.**  $MA(1)$  как  $MA(\infty)$ .

**Пример 3.**  $MA(1)$  как  $MA(1)$  с другими коэффициентами.

### Частная корреляция

Обычная корреляция

$$\text{Corr}(L, R) = \frac{\text{Var}(L, R)}{\sqrt{\text{Var}(L) \text{Var}(R)}}$$

измеряет силу линейной связи между величинами  $L$  и  $R$ .

Частная корреляция

$$\text{pCorr}(L, R; M_1, M_2, M_3)$$

измеряет силу линейной связи между величинами  $L$  и  $R$ , если обрвать связи проходящие через величины  $M_1, M_2, M_3$ . Величин, связи через которые элиминируются, не обязательно три. Их может быть любое количество.

**Определение 6.** Частная корреляция

$$\text{pCorr}(L, R; M_1, M_2, M_3) = \text{Corr}(\tilde{L}, \tilde{R}),$$

где  $\tilde{L}$  — величина  $L$ , «очищенная» от  $M_i$ -х, а  $\tilde{R}$  — величина  $R$ , очищенная от  $M_i$ -х. «Очищенные» версии строятся по формулам

$$\tilde{L} = L - (\alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3), \quad \tilde{R} = R - (\beta_1 M_1 + \beta_2 M_2 + \beta_3 M_3),$$

так, чтобы у очищенных величин была нулевая ковариация с любой  $M_i$ ,

$$\text{Var}(L, M_i) = 0, \quad \text{Var}(R, M_i) = 0.$$

**Теорема 1.** У стационарного процесса  $(y_t)$  частная корреляция

$$\phi_{kk} = \text{pCorr}(y_t, y_{t+k}; y_{t+1}, \dots, y_{t+k-1})$$

может быть найдена из разложения

$$y_t = \alpha + \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \dots + \phi_{kk} y_{t-k} + w_t,$$

где ковариации  $\text{Var}(w_t, y_{t-1}), \dots, \text{Var}(w_t, y_{t-k})$  равны нулю.

Из системы уравнений

$$\begin{cases} \text{Var}(w_t, y_{t-1}) = 0 \\ \text{Var}(w_t, y_{t-2}) = 0 \\ \vdots \text{Var}(w_t, y_{t-k}) = 0 \end{cases}$$

можно найти все частные корреляции  $\phi_{k1}, \dots, \phi_{kk}$ . Система уравнений называется системой Юла-Волкера (Yule-Walker equations).

Наиболее легко интерпретируется и чаще используется частная корреляция  $\phi_{kk}$ . Однако можно проинтерпретировать и другие, например,

$$\phi_{42} = \text{pCorr}(y_t, y_{t-2}; y_{t-1}, y_{t-3}, y_{t-4}).$$

Применив правило Крамера можно получить готовую формулу

$$\phi_{kk} = \frac{\det A}{\det B},$$

где в матрице  $B$  элемент  $b_{ij} = \text{Corr}(y_i, y_j)$ , а матрица  $A$  отличается от матрицы  $B$  тем, что вместо последнего столбца написали  $a_{in} = \text{Corr}(y_t, y_{t+i})$ .

Если взять ожидание от левой и правой части разложения  $y_t$ , то можно найти и константу  $\alpha$ .

Разложение

$$y_t = \alpha + \phi_{k1}y_{t-1} + \phi_{k2}y_{t-2} + \dots + \phi_{kk}y_{t-k} + w_t,$$

из уравнения для поиска частной корреляции можно использовать также и для построения наилучших линейных прогнозов. А именно, если мы знаем последние  $k$  значений нашего ряда  $(y_t)$ , то наилучший линейный прогноз имеет вид

$$\hat{y}_{t+1|t} = \alpha + \phi_{k1}y_t + \phi_{k2}y_{t-1} + \dots + \phi_{kk}y_{t-k+1}.$$

Заметим, что для процесса  $(y_t)$  с совместным нормальным распределением наилучшие линейные прогнозы будут просто наилучшими.

### *Строго про сходимости*

#### *Оператор лага*

**Определение 7.** Оператор лага  $L$  переводит случайный процесс  $(y_t)$  в случайный процесс  $(\tilde{y}_t)$  по формуле  $\tilde{y}_t = y_{t-1}$ .

Строго говоря, надо использовать обозначение  $L((y_t))$ , потому что случайный процесс мы обозначаем как  $(y_t)$ , а оператор применяется именно к случайному процессу. Однако на практике пишут  $\tilde{y}_t = Ly_t$  и все понимают, что  $\tilde{y}_t = y_{t-1}$ .

Можно построить пример-ловушку, основанный на этой тонкой разнице. Рассмотрим два процесса, связанных соотношением  $x_t = y_{-t}$ . Что такое  $Lx_5$ ? С одной стороны,  $Lx_5 = x_4$ . С другой стороны,  $x_5 = y_{-5}$ , это одна и та же случайная величина, следовательно,  $Lx_5 = Ly_{-5} = y_{-6} = x_6$ . Противоречия не возникает, если чётко осознавать, что оператор лага  $L$  применяют не к отдельно взятой случайной величине, а случайному процессу в целом, то есть, к последовательности случайных величин.

## Разностные уравнения и ARMA

**Определение 8.** Ряд  $(y_t)$  называется процессом авторегрессии порядка  $p$  (autoregression of order  $p$ ) относительно белого шума  $(u_t)$ , если выполнено два условия: ряд  $(y_t)$  представим в виде

$$y_t - \mu = \beta_1(y_{t-1} - \mu) + \beta_2(y_{t-2} - \mu) + \dots + \beta_p(y_{t-p} - \mu) + u_t, \quad (1)$$

где последний коэффициент  $\beta_p \neq 0$ . ряд  $(y_t)$  является процессом  $MA(\infty)$  относительно белого шума  $(u_t)$ :

$$y_t = \mu + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots$$

Обозначаем такие процессы мы так:  $y_t \sim AR(p)$ .

Следует отметить, что во многих учебниках не дано корректного определения  $AR(p)$  процесса. Очень часто авторы ограничиваются в определении уравнением

$$y_t - \mu = \beta_1(y_{t-1} - \mu) + \beta_2(y_{t-2} - \mu) + \dots + \beta_p(y_{t-p} - \mu) + u_t.$$

в остальном, возможно, прекрасных

Проблема состоит в том, что этому уравнению удовлетворяет бесконечное количество случайных процессов. Разберём на примере.

**Пример 4.** Рассмотрим уравнение  $y_t = 0.5y_{t-1} + u_t$ . Легко найти нестационарное решение. Выбираем любой момент времени, скажем  $t = 2$ , выбираем любое число, например, 100, и дополняем уравнение начальным условием  $y_2 = 100$ .

Все остальные  $y_t$  находятся однозначно. В частности  $y_3 = 50 + u_3$ ,  $y_4 = 25 + 0.5u_3 + u_4$ , .... Можно раскручивать последовательность и в прошлое, переписав рекуррентную формулу в виде  $y_{t-1} = 2y_t - 2u_t$ . Отсюда  $y_1 = 200 - 2u_2$ ,  $y_0 = 400 - 4u_2 - 2u_1$ , ....

Конечно, получающийся процесс нестационарен,  $E(y_2) = 100$ ,  $E(y_3) = 50$ .

**Пример 5.** Рассмотрим уравнение  $y_t = 2y_{t-1} + u_t$  и покажем, что у него *есть* стационарное решение.

Некоторые авторы, например, замечательный Аад ван дер Ваарт Van der Vaart 2010, определяют  $AR(p)$  процесс как любое решение уравнения 1. Мы пошли по другому пути, чтобы сделать определение ближе к формулировке, фактически используемой в статистических пакетах.

**Теорема 2.** Рассмотрим уравнение на процесс  $(y_t)$

$$y_t - \mu = \beta_1(y_{t-1} - \mu) + \beta_2(y_{t-2} - \mu) + \dots + \beta_p(y_{t-p} - \mu) + u_t,$$

где  $\beta_p \neq 0$ ,  $p \geq 1$  и  $(u_t)$  — белый шум.

Уравнение имеет

бесконечное количество нестационарных решений.

ровно одно стационарное решение вида  $MA(\infty)$  относительно шума  $(u_t)$ , если и только если все корни характеристического уравнения  $|\lambda| < 1$ .

ровно одно стационарное решение, если и только если все корни характеристического уравнения  $|\lambda| \neq 1$ .

Очень часто эту теорему просто и ошибочно формулируют как «процесс  $AR(p)$  стационарен, если все корни характеристического уравнения по модулю меньше 1». Как мы видели, у рекуррентного уравнения  $y_t = 0.5y_{t-1} + u_t$  будет множество нестационарных решений, а у рекуррентного уравнения  $y_t = 2y_{t-1} + u_t$  существует стационарное решение.

### Стационарный бутстрэп

Представим себе, что у нас есть ряд  $y_1, \dots, y_T$ , и мы хотим построить доверительный интервал для  $\rho = \text{Corr}(y_t, y_{t-1})$  с помощью бутстрэпа.

Если использовать обычный бутстрэп, который из исходной выборки  $(y_t)$  много раз делает случайную выборку с повторениями, то структура временного ряда будет разрушаться при создании бутстрэп-выборок, и оценка корреляции по бутстрэп-выборкам будет каждый раз примерно нулевой.

Алгоритм стационарного бутстрэпа пытается решить эту проблему. На входе у нас временной ряд  $y_1, \dots, y_T$ . На выходе мы хотим получить бутстрэп копию этого ряда той же длины  $y_1^*, \dots, y_T^*$ .

- а) Выберем параметр вероятности  $p$ . О правилах выбора чуть позже.
- б) Выберем случайный момент времени  $s \in \{1, \dots, T\}$  и запишем  $y_s$  очередным элементом в бутстрэп копию.
- в) С вероятностью  $p$  вернемся к шагу 2, с вероятностью  $1 - p$  пойдём дальше.
- г) Увеличим  $s$  на 1, запишем  $y_s$  очередным элементом в бутстрэп копию и перейдем к подкидыванию монетки на шаге 3.

Алгоритм продолжается до тех пор, пока не наберем  $T$  наблюдений в бутстрэп-копию ряда.

Теперь мы можем построить бутстрэп-доверительный интервал для корреляции. Например, с помощью перцентильного бутстрэпа.

По исходному ряду создаем 10000 бутстрэп-копий ряда. По каждой бутстрэп-копии считаем оценку корреляции. Удаляем по 2.5% самых больших и самых маленьких оценок корреляции. Полученные края и будут границами доверительного интервала.

Про выбор  $p$ .

...

## ETS

### Сглаживание ряда

При сглаживании ряда мы из исходного ряда  $(y_t)$  получаем новый ряд  $(\tilde{y}_t)$  с меньшей изменчивостью. Количество наблюдений при этом может как немного поменаться, так и сохраниться, в зависимости от конкретного алгоритма.

### Скользящее среднее

**Определение 9.** Взятие скользящего среднего с шириной окна  $h = 3$  — алгоритм сглаживания с формулой

$$\tilde{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}$$

или

$$\tilde{y}_t = \frac{y_{t-2} + y_{t-1} + y_t}{3}.$$

Для краткости можно использовать обозначение  $\tilde{y} = MA(y)$ .

Как выглядит скользящее среднее с другой нечётной шириной окна читатель может попробовать догадаться сам, к примеру, выписав формулу для скользящего среднего с шириной окна  $h = 5$ . При взятии скользящего среднего мы либо теряем наблюдения в начале и в конце ряда, либо нам нужно как-то адаптировать эту формулу для первого и последнего наблюдения. Например, для последнего наблюдения можно взять

$$\tilde{y}_T = \frac{y_{T-1} + y_T}{2}.$$

Стоит отметить и то, что скользящее среднее с формулой

$$\tilde{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}.$$

немного «подглядывает в будущее». Действительно, в формулу для  $\tilde{y}_t$  входит  $y_{t+1}$ . В некоторых случаях это может завышать оценку качества прогнозов.

## LOESS

Обычная регрессия (ordinary least squares) проводит одну линию  $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_t$ . В роли  $x_t$  может быть просто само время  $t$ .

Вспомним целевую функцию обычной регрессии

$$Q(\hat{\beta}_1, \hat{\beta}_2) = \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \sum_{t=1}^T (y_t - (\hat{\beta}_1 + \hat{\beta}_2 x_t))^2.$$



В результате минимизации

$$Q(\hat{\beta}_1, \hat{\beta}_2) \rightarrow \min_{\hat{\beta}_1, \hat{\beta}_2} Q(\hat{\beta}_1, \hat{\beta}_2)$$

получается единственное значение  $\hat{\beta}_1$  и  $\hat{\beta}_2$ .

LOESS = LOcal regrESSion = ЛОкальная регрЕССия

LOESS проводит свою линию регрессии для каждого  $x$ . Целевая функция теперь зависит от абсциссы точки  $x$ , в которой мы строим регрессию,

$$Q(\hat{\beta}_1, \hat{\beta}_2) = \sum_{t=1}^T K(x_t, x)(y_t - \hat{y}_t)^2 = \sum_{t=1}^T K(x_t, x)(y_t - (\hat{\beta}_1 + \hat{\beta}_2 x_t))^2.$$

Функция весов  $K(x_t, x)$  должна давать большой положительный вес точкам  $x_t$  рядом с точкой  $x$  и маленький положительный, или даже нулевой, вес точкам  $x_t$  далеко от точки  $x$ .

Например, в качестве функции весов  $K(x_t, x)$  можно использовать

$$K(x_t, x) = \exp\left(-\frac{(x_t - x)^2}{h^2}\right).$$

При  $h \rightarrow \infty$  мы получим обычные оценки метода наименьших квадратов  $\hat{\beta}_1(x) = \hat{\beta}_1^{\text{OLS}}$ ,  $\hat{\beta}_2(x) = \hat{\beta}_2^{\text{OLS}}$ .

Функция весов может быть и такой

$$K(x_t, x) = \begin{cases} 1, & \text{если } x_t \text{ — это один из пяти ближайших соседей } x, \\ 0, & \text{иначе} \end{cases}.$$

Оптимизируем мы по-прежнему по двум переменным,

$$Q(\hat{\beta}_1, \hat{\beta}_2, x) \rightarrow \min_{\hat{\beta}_1, \hat{\beta}_2} Q(\hat{\beta}_1, \hat{\beta}_2, x).$$

Задача оптимизации — выпуклая, есть решение в явном виде. Только теперь получаются оптимальные коэффициенты, зависящие от точки  $x$ ,  $\hat{\beta}_1(x)$  и  $\hat{\beta}_2(x)$ .

Для получения сглаженного значения  $\tilde{y}_t$  мы берём  $x_t = t$  и

$$\tilde{y}_t = \hat{\beta}_1(t) + \hat{\beta}_2(t)t.$$

## Выделение сезонности

### STL

Изложим упрощённый вариант STL-алгоритма для месячных данных без выбросов.

а) Положим  $\text{trend}_t = 0$  и  $\text{season}_t = 0$ .

б) Детрендируем исходный ряд,

$$D_t = y_t - \text{trend}_t.$$

в) Разрежем детрендированный ряд  $(D_t)$  на двенадцать подрядов по месяцам.

$$(D_t) \rightarrow (D_t^{jan}), (D_t^{feb}), \dots, (D_t^{dec}).$$

г) Сгладим каждый подряд с помощью LOESS:

$$C^{jan} = LOESS(D^{jan}), \dots, C^{dec} = LOESS(D^{dec})$$

д) Соберём двенадцать подрядов в один ряд

$$(C_t^{jan}), (C_t^{feb}), \dots, (C_t^{dec}) \rightarrow (C_t)$$

е) Сильно сгладим собранный ряд

$$L = LOESS(MA(MA(C))).$$

ж) Обновим сезонную составляющую

$$\text{season}_t = C_t - L_t.$$

з) Обновим тренд

$$\text{trend}_t = y_t - \text{season}_t.$$

Далее перейдём к шагу 2 и пройдем шаги 2-8 ещё раз.

*MSTL*

*Байесовские модели*

*DLT*

Вспомним ETS(A, Ad, A) модель в структурной форме.

$$\begin{cases} y_t = (\ell_{t-1} + \phi b_{t-1}) + s_{t-12} + u_t \\ s_t = s_{t-12} + \gamma u_t \\ \ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha u_t \\ b_t = \phi b_{t-1} + \beta u_t \\ u_t \sim \text{Normal}(\text{loc} = 0, \text{scale} = \sigma). \end{cases}$$

Стандартной ссылкой по ETS моделям является Hyndman и Athanasopoulos 2018.

Чтобы перейти к DLT модели сделаем ряд обобщений:

- Добавим в наблюдаемый процесс  $y_t$  регрессионную составляющую:

$$r_t = \beta_1 x_{t1} + \beta_2 x_{t2}.$$

- Добавим в наблюдаемый процесс глобальный тренд, например, линейный:

$$g_t = \delta_1 + \delta_2 t.$$

- Перейдем от нормального распределения ошибки к распределению Стьюдента,

$$u_t \sim \text{Student}(\text{df} = \nu, \text{loc} = 0, \text{scale} = \sigma).$$

Кроме того, в описании DLT модели по сравнению с ETS моделью почему-то сдвинут индекс у сезонной составляющей. То есть величина, называемая  $s_t$  у Хиндмана в Hyndman и Athanasopoulos 2018, в статье Ng и др. 2020 названа  $s_{t+12}$ . Поэтому уравнение на сезонную составляющую принимает вид

$$s_{t+12} = s_t + \gamma u_t.$$

С учётом новых составляющих и сдвига индекса у сезонности уравнение на наблюдаемый  $y_t$  примет вид

$$y_t = g_t + (\ell_{t-1} + \phi b_{t-1}) + s_t + r_t + u_t.$$

### DLT модель

Наблюдаемый ряд  $y_t$  раскладывается в сумму составляющих: глобальный тренд, локальное отклонение от глобального тренда, сезонная составляющая, регрессионная составляющая, ошибка.

$$y_t = g_t + (\ell_{t-1} + \phi b_{t-1}) + s_t + r_t + u_t$$

Глобальный тренд  $g_t$  может быть задан по-разному. Например, линейно

$$g_t = \gamma_1 + \gamma_2 t.$$

Сезонная составляющая плавно меняется во времени,

$$s_{t+12} = s_t + \gamma u_t.$$

Скорость локального тренда плавно меняется,

$$b_t = \phi b_{t-1} + \beta u_t.$$

Локальный тренд

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha u_t.$$

Регрессионная составляющая на примере двух регрессоров,

$$r_t = \beta_1 x_{t1} + \beta_2 x_{t2}.$$

Ошибка,

$$u_t \sim \text{Student}(\text{df} = \nu, \text{loc} = 0, \text{scale} = \sigma).$$

Одной системой,

$$\begin{cases} y_t = g_t + (\ell_{t-1} + \phi b_{t-1}) + s_t + r_t + u_t \\ g_t = \gamma_1 + \gamma_2 t \\ s_{t+12} = s_t + \gamma u_t \\ r_t = \beta_1 x_{t1} + \beta_2 x_{t2} \\ \ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha u_t \\ b_t = \phi b_{t-1} + \beta u_t \\ u_t \sim \text{Student}(\text{df} = \nu, \text{loc} = 0, \text{scale} = \sigma). \end{cases}$$

Используя формулу для  $\ell_t$  можно записать  $y_t$  также в виде

$$y_t = g_t + \ell_t + s_t + r_t + (1 - \alpha)u_t.$$

### Рекуррентные соотношения

Можно элегантно отказаться от  $u_t$  в уравнениях на  $s_{t+12}$ ,  $\ell_t$  и  $b_t$ . Это полезно для описания модели на вероятностных языках программирования, будь то stan, numrug или что-то ещё.

Выразим ошибку  $u_t$  из формулы для локального тренда и подставим в формулу для скорости роста локального тренда :

$$b_t = \phi b_{t-1} + \frac{\beta}{\alpha}(\ell_t - \ell_{t-1} - \phi b_{t-1}).$$

Перегруппируем и увидим, что скорость роста локального тренда  $b_t$  является средневзвешенным,

$$b_t = \frac{\beta}{\alpha}(\ell_t - \ell_{t-1}) + \left(1 - \frac{\beta}{\alpha}\right) \phi b_{t-1}.$$

Можно определить  $\rho_b = \frac{\beta}{\alpha}$ , и тогда

$$b_t = \rho_b(\ell_t - \ell_{t-1}) + (1 - \rho_b)\phi b_{t-1}.$$

В коде пакета orbit на stan соответствующая строка имеет вид

$$b[t] = \text{slp\_sm} * (l[t] - l[t-1]) + (1 - \text{slp\_sm}) * \text{DAMPED\_FACTOR} * b[t-1];$$

Теперь выразим ошибку  $u_t$  из и подставим в .

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(y_t - g_t - \ell_{t-1} - \phi b_{t-1} - s_t - r_t).$$

Перегруппируем и снова получаем вид средневзвешенного.

$$\ell_t = \alpha(y_t - g_t - s_t - r_t) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}).$$

Смотрим на исходный код модели в stan,

$$\begin{aligned} \text{lt\_sum}[t] &= l[t-1] + \text{DAMPED\_FACTOR} * b[t-1]; \\ l[t] &= \text{lev\_sm} * (\text{RESPONSE}[t] - \text{gt\_sum}[t] - s_t - r[t]) + (1 - \text{lev\_sm}) * \text{lt\_sum}[t]; \end{aligned}$$

На этот раз выразим ошибку  $u_t$  из и подставим в .

$$s_{t+12} = s_t + \frac{\gamma}{1 - \alpha}(y_t - g_t - \ell_t - s_t - r_t).$$

Перегруппируем и получаем вид средневзвешенного,

$$s_{t+12} = \frac{\gamma}{1 - \alpha}(y_t - g_t - \ell_t - r_t) + \left(1 - \frac{\gamma}{1 - \alpha}\right) s_t.$$

При обозначении  $\rho_s = \frac{\gamma}{1 - \alpha}$  получаем

$$s_{t+12} = \rho_s(y_t - g_t - \ell_t - r_t) + (1 - \rho_s)s_t.$$

Соответствующий фрагмент кода в stan,

$$\begin{aligned} s[t + \text{SEASONALITY}] &= \text{sea\_sm} * (\text{RESPONSE}[t] - \text{gt\_sum}[t] - l[t] \\ &- r[t]) + (1 - \text{sea\_sm}) * s_t; \end{aligned}$$

Замечаем, что в статье Ng и др. 2020 в описании DLT модели есть пара описок. Пропущено  $\phi$  перед  $b_{t-1}$  в рекуррентной формуле для  $\ell_t$ . Пропущено  $g_t$  в рекуррентной формуле для  $s_{t+m}$ .

*Начальные условия*

$$b_1 = 0$$

$$g_1 =$$

$$r_1 =$$

$$s_{12} = -(s_1 + s_2 + \dots + s_{11})$$

$$\ell_1 = y_1 - g_1 - s_1 - r_1$$

*Априорные распределения*

При  $t \in \{1, 2, \dots, 11\}$ ,

$$s_t \sim \text{Normal}(\text{loc} = 0, \text{scale} = \sigma?)$$

```
for (i in 1:(SEASONALITY - 1))
  init_sea[i] ~ normal(0, SEASONALITY_SD);
```

$$\beta_j \sim \text{Normal}(\text{loc} = \mu_j, \text{scale} = \sigma_j),$$

где  $\mu_j, \sigma_j$  — гиперпараметры, по умолчанию равные  $\mu_j = 0$  и  $\sigma_j = 1$ .

$$\sigma \sim \text{HalfCauchy}(\text{loc} = 0, \text{scale} = \gamma_0),$$

где  $\gamma_0 \dots$

*Тесты на прогнозную силу**Тест Диболда-Мариано*

Предпосылки:

Два прогноза,  $\hat{y}_t^A$  и  $\hat{y}_t^B$ . Разница произвольных метрик качества,

$$d_t = (\hat{y}_t^A - y_t)^2 - (\hat{y}_t^B - y_t)^2.$$

Процесс  $(d_t)$  стационарный. Другими словами  $E(d_t) = \mu$ ,  $\text{Var}(d_t, d_{t-k}) = \gamma_k$ , в частности,  $\text{Var}(d_t) = \gamma_0$ .

Гипотезы:

$$H_0: E(d_t) = 0;$$

$$H_a: E(d_t) \neq 0;$$

Тестовая статистика при верной  $H_0$ :

$$DM = \frac{\bar{d} - 0}{se(\bar{d})} \rightarrow \mathcal{N}(0; 1)$$

Трудность возникает только в оценке  $se(\bar{d})$ , так как значения  $d_t$  коррелированы.

Как правило оценивают регрессию вектора  $d_t$  на константу и используют робастную стандартную ошибку  $se_{HAC}$ .

$$\hat{d}_t = \hat{\beta}_1, \quad DM = \frac{\hat{\beta}_1 - 0}{se_{HAC}(\hat{\beta}_1)}.$$

В качестве альтернативного подхода можно дополнительно предположить, что  $(d_t)$  описывается стационарным  $ARMA(p, q)$  процессом с небольшими  $p$  и  $q$  и рассчитать  $se(\bar{d})$  в рамках этого предположения.

*RC и SPA тесты*

RC (Reality Check) тест Уайта и SPA (Superior Predictive Ability) тест Хансена обобщают тест Диболда-Мариано на случай сравнения множества прогнозов против одного эталонного.

Для обоих тестов используется стационарный бутстрэп.

*Решения**Источники мудрости*

Источники мудрости, кои автор подборки постарался не замутить. Смело направляйте к ним верблюдов своего любопытства!

*Список литературы*

- Hyndman, Rob J и George Athanasopoulos (2018). *Forecasting: principles and practice*. OTexts. URL: <https://otexts.com/fpp3/>.
- Ng, Edwin и др. (2020). «Orbit: probabilistic forecast with exponential smoothing». В: *arXiv preprint arXiv:2004.08492*. URL: <https://arxiv.org/abs/2004.08492>.
- Van der Vaart, Aad W (2010). «Time series». В: *VU University Amsterdam, lecture notes*. URL: <https://staff.fnwi.uva.nl/p.j.c.spreij/onderwijs/master/aadtimeseries2010.pdf>. try vpn in Netherlands.