Авторегрессионные модели прогнозирования

Д. А. Ивахненко. Методы прогнозирования, СПбГЭУ 2021 г.

Литература

G. E. P. Box, D. R. Cox. An analysis of transformations, Journal of the Royal Statistical Society, Series B, 26, 211-252 (1964)

---

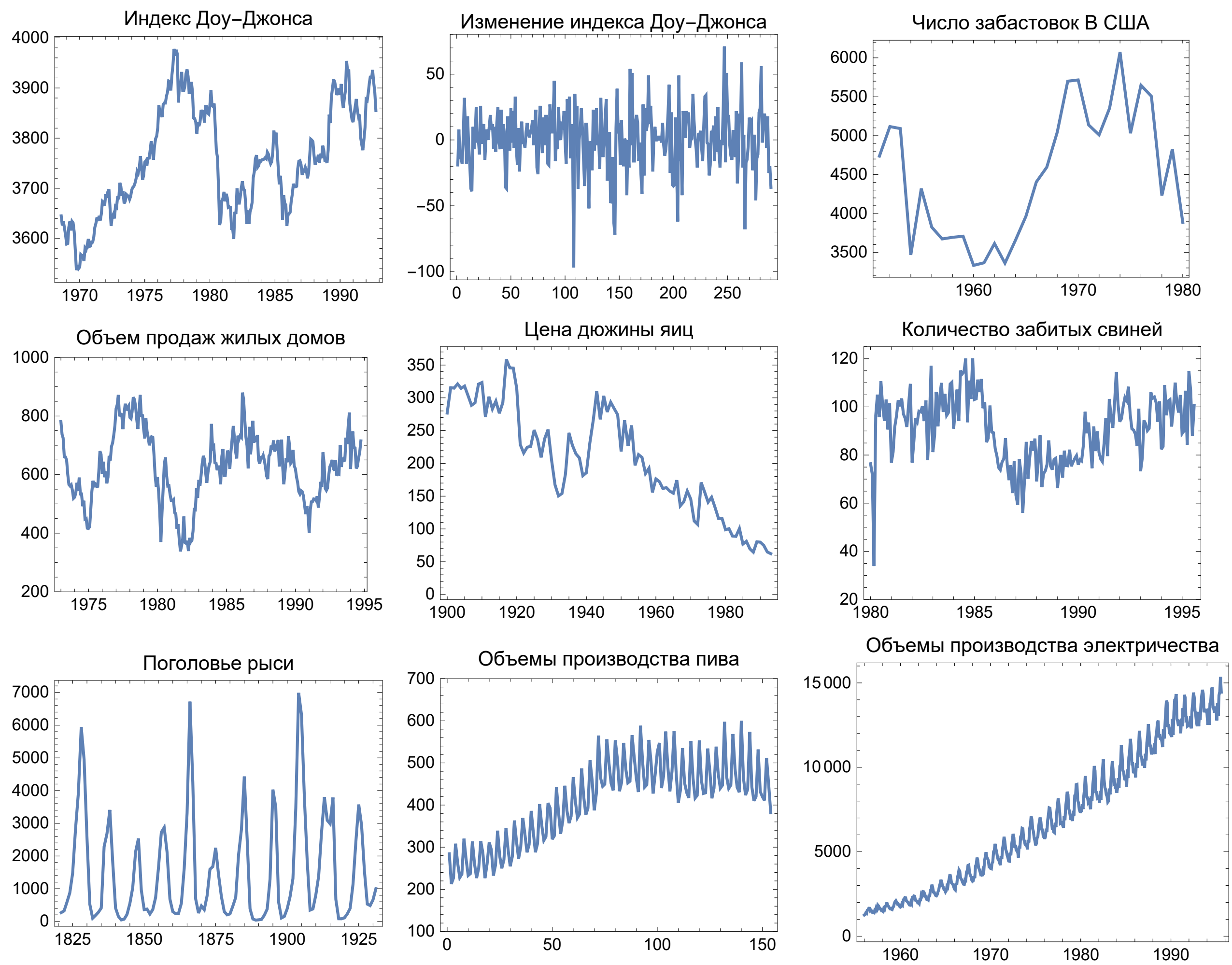
1. Понятие стационарности

Временной ряд $y_1, \ldots, y_T$ называется **стационарным**, если $\forall\, k$ (ширины окна) распределение $y_t, \ldots, y_{t+k}$ не зависит от t, т.е. его свойства не зависят от времени.

Из этого определения следует, что ряды, в которых присутствует тренд, являются нестационарными: в зависимости от расположения окна изменяется средний уровень ряда. Кроме того, нестационарны ряды с сезонностью: если ширина окна меньше сезонного периода, то распределение ряда будет разным, в зависимости от положения окна. При этом интересно, что ряды, в которых есть непериодические циклы, не обязательно являются нестационарными, поскольку нельзя заранее предсказать положение максимумов и минимумов этого ряда.

Упражнение

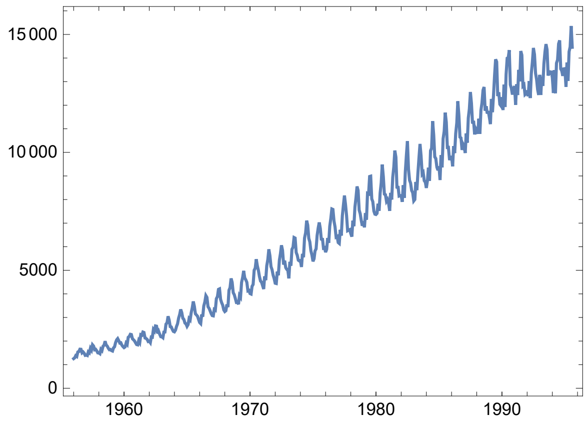
Какие из представленных ниже рядов являются стационарными?

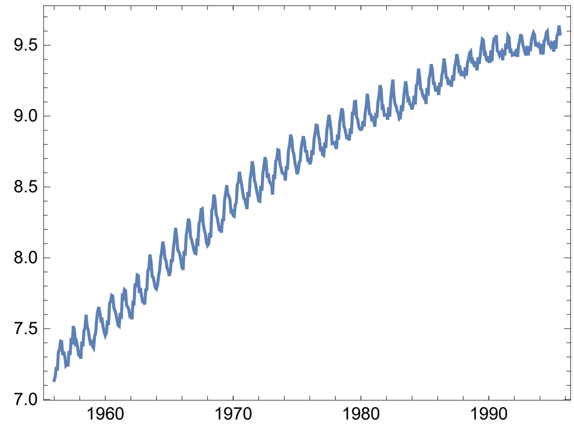


Гипотезу о стационарности можно проверить с помощью критерия Дики-Фуллера, который будет рассмотрен далее.

1.1. Стабилизация дисперсии

Если во временном ряде монотонно по времени изменяется дисперсия, применяется специальное преобразование, стабилизирующее дисперсию. Часто в качестве такого преобразования выступает логарифмирование. В результате логарифмирования ряда производства электричества в Австралии размах колебаний в начале и конце ряда становится очень похожим, и дисперсия стабилизируется.



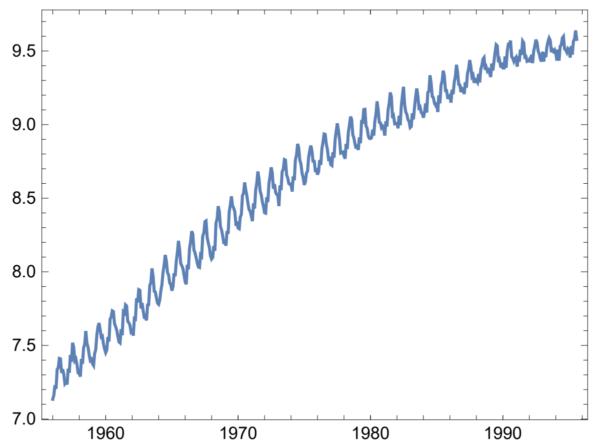


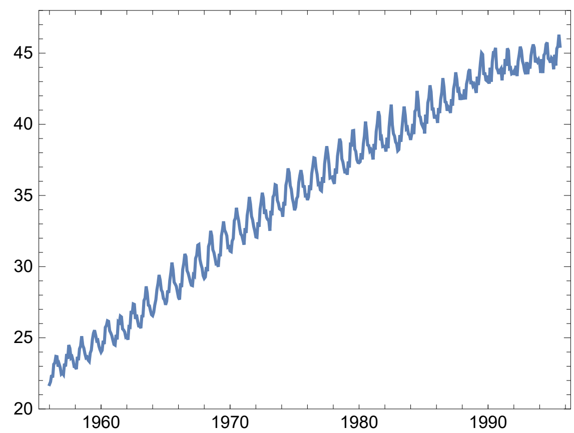
Логарифмирование принадлежит к параметрическому семейству преобразований Бокса-Кокса. В случае, когда значения ряда $y > 0$, преобразование Бокса-Кокса имеет вид:

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log y, & \lambda = 0. \end{cases}$$

Заметим, что $y^\lambda = e^{\lambda \log(y)} = 1 + \lambda \log(y) + O((\lambda \log(y))^2)$. Тогда $y^{(\lambda)} = \log(y)$ в случае, когда λ бесконечно мало.

Параметр λ определяет, как именно будет преобразован ряд: $\lambda = 0$ – логарифмирование, $\lambda = 1$ – тождественное преобразование ряда, при других значениях λ – степенное преобразование. Значение параметра можно подбирать так, чтобы дисперсия была как можно более стабильной во времени. Так, для ряда по данным производства электричества в Австралии оптимальное значение $\lambda = 0.27$.





Параметр λ выбирается методом максимального правдоподобия. Преобразование Бокса-Кокса относится к семейству степенных преобразований вида:

$$
y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda(\mathrm{gm}(y))^{\lambda-1}}, & \lambda \neq 0, \\ \mathrm{gm}(y)\log y, & \lambda = 0, \end{cases}
$$

где $\mathrm{gm}(y) = \left(\prod_{i=1}^{T} y_i\right)^{\frac{1}{T}} = \sqrt[T]{y_1 y_2 \dots y_T}$ – среднее геометрическое ряда.

Бокс и Кокс в своей статье включили среднее геометрическое в преобразование, связав плотность распределения исходного ряда с плотностью преобразованного следующим соотношением:

$$
J(\lambda;\, y_1,\, y_2,\, \dots,\, y_T) = \prod_{i=1}^{T} \left| \frac{\partial y_i^{(\lambda)}}{\partial y_i} \right| = \prod_{i=1}^{T} y_i^{\lambda-1} = \mathrm{gm}(y)^{T(\lambda-1)},
$$

$$
f(y_1,\, \dots,\, y_T) = f_{(\lambda)}\left(y_1^{(\lambda)},\, \dots,\, y_T^{(\lambda)}\right) J(\lambda;\, y_1,\, y_2,\, \dots,\, y_T).
$$

Из предположения, что значения ряда $y_i^{(\lambda)}$ $(i = 1,\, \dots,\, T)$ распределены нормально с математическим ожиданием $\overline{y}^{(\lambda)}$ и постоянной дисперсией σ^2, оценка параметра λ может быть получена путем максимизации логарифма правдоподобия:

$$
L_{\max}(\lambda) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(- \frac{\left(y_i^{(\lambda)} - \overline{y}^{(\lambda)}\right)^2}{2\sigma^2} \right) J(\lambda;\, y),
$$

$$
\log(L_{\max}(\lambda)) = -\frac{T}{2} \log\left(\frac{\sum_i \left(y_i^{(\lambda)} - \overline{y}^{(\lambda)}\right)^2}{T} \right) + (\lambda - 1) \sum_i \log(y_i).
$$

Если ряд содержит отрицательные значения, то можно переписать правила преобразования следующим образом:

$$
y^{(\lambda)} = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1}, & \lambda_1 \neq 0, \\ \log\,(y + \lambda_2), & \lambda_1 = 0, \end{cases}
$$

где $y > -\lambda_2$.

1.2. Дифференцирование

Еще один важный трюк, который позволяет сделать ряд стационарным, – это дифференцирование, переход к попарным разностям соседних значений:

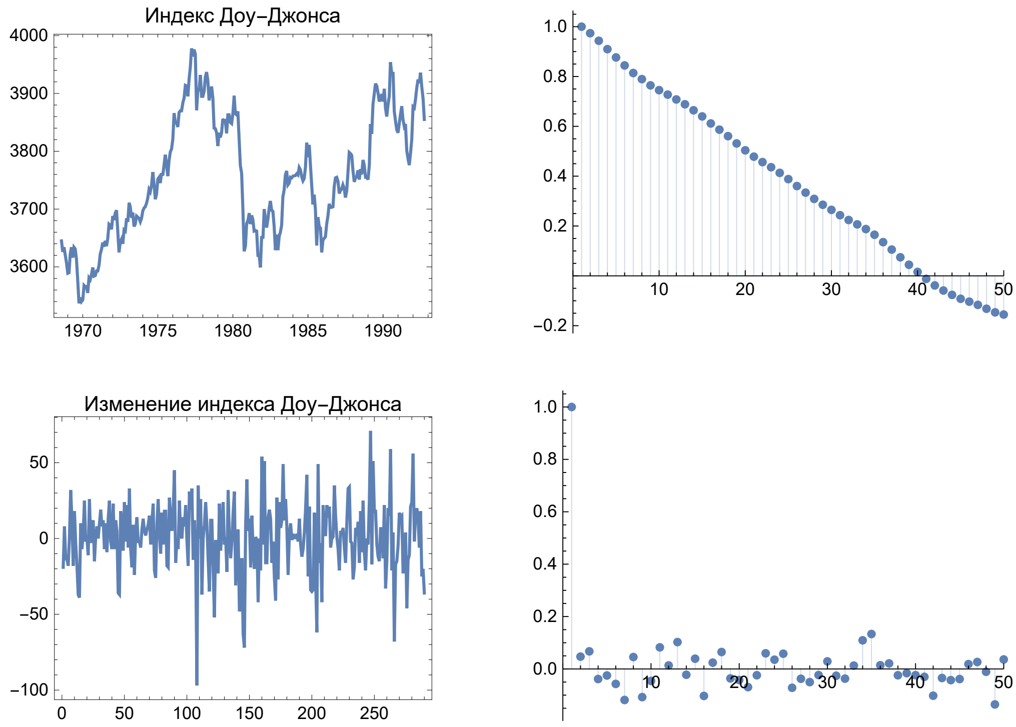
$$
y'_t = y_t - y_{t-1}.
$$

Для нестационарного ряда часто оказывается, что получаемый после дифференцирования ряд является стационарным. Такая операция позволяет стабилизировать среднее значение ряда и избавиться от тренда, а иногда даже от сезонности. Кроме того, дифференцирование можно применять неоднократно: от ряда первых разностей, продифференцировав его, можно прийти к ряду вторых разностей, и т. д. Длина ряда при этом каждый раз будет немного сокращаться, но при этом он будет стационарным.

Также может применяться сезонное дифференцирование ряда, переход к попарным разностям значений в соседних сезонах. Если длина периода сезона составляет s, то новый ряд задается разностями:

$$
y'_t = y_t - y_{t-s}.
$$

Сезонное и обычное дифференцирование могут применяться к ряду в любом порядке. Однако если у ряда есть ярко выраженный сезонный профиль, то рекомендуется начинать с сезонного дифференцирования, уже после такого преобразования может оказаться, что ряд стационарен.



На верхних графиках показаны ряд значений индекса Доу-Джонса и его автокорреляционная функция. Видно, что этот ряд нестационарен – имеется ярко выраженный тренд. От тренда удается полностью избавиться, продифференцировав ряд.

Таким образом, для приведения временного ряда к стационарному первым делом необходимо стабилизировать дисперсию, то есть применить преобразование Бокса-Кокса, затем, при наличии ярко выраженной сезонности провести сезонное дифференцирование с лагом, равным сезонному периоду. При необходимости провести обычное дифференцирование.

1.3. Обратное преобразование

Переход от полученного стационарного временного ряда к исходному может быть выполнен по следующему правилу:

$$y'_t = y_t - y_{t-k},$$
$$y_t = y'_t + y_{t-k},$$

где y_t – исходный ряд, а k – лаг дифференцирования.

2. Модели класса ARMA

2.1. ARMA

Модель ARMA – сумма авторегрессионной модели порядка p (AR(p)) и модели скользящего среднего порядка q (MA(q)):

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}.$$

Теорема Вольда: любой стационарный ряд может быть описан моделью ARMA(p, q).

2.2. ARIMA

В основе моделей класса ARIMA лежат идеи о том, что нестационарный ряд можно сделать стационарным при помощи дифференцирования, а любой стационарный ряд может быть описан моделью ARMA(p, q).

Модель авторегресии интегрированного скользящего среднего **ARIMA(p, d, q)** – это модель ARMA(p, q) для d раз продифференцированного ряда.

2.3. SARMA

Пусть ряд имеет сезонный период длины S. Возьмем модель ARMA(p, q):

$$y_t = \alpha + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \ldots + \phi_p\, y_{t-p} + \epsilon_t + \theta_1\, \epsilon_{t-1} + \theta_2\, \epsilon_{t-2} + \ldots + \theta_p\, \epsilon_{t-q},$$

добавим P авторегрессионных компонент:

$$+\phi_S\, y_{t-S} + \phi_{2\,S}\, y_{t-2\,S} + \ldots + \phi_{P\,S}\, y_{t-P\,S}$$

и Q компонент скользящего среднего:

$$+\theta_S\, \epsilon_{t-S} + \theta_{2\,S}\, \epsilon_{t-2\,S} + \ldots + \theta_{Q\,S}\, \epsilon_{t-Q\,S}.$$

Результат – модель SARMA(p, q) $\times$ (P, Q).

2.4. SARIMA

Модель SARIMA(p, d, q) $\times$ (P, D, Q) – модель SARMA (p, q) $\times$ (P, Q) для ряда, к которому d раз было применено обычное дифференцирование и D раз – сезонное. Такую модель часто называют просто ARIMA: первая буква не пишется, но подразумевается, что сезонная компонента тоже может быть.

2.5. Построение прогноза

Теперь необходимо разобраться, как на основании настроенной модели ARIMA правильно строить прогноз. Пусть модель построена, определены значения всех неизвестных параметров, получены их оценки $\hat{\alpha}$; $\hat{\varphi}$; $\hat{\theta}$, которые записаны в уравнении:

$$y_t = \hat{\alpha} + \hat{\phi}_1\, y_{t-1} + \ldots + \hat{\phi}_p\, y_{t-p} + \epsilon_t + \hat{\theta}_1\, \epsilon_{t-1} + \ldots + \hat{\theta}_p\, \epsilon_{t-q}.$$

Чтобы построить прогноз на момент времени $T+1$, нужно в этом уравнении заменить все индексы t на $T+1$:

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1\, y_T + \ldots + \hat{\phi}_p\, y_{T+1-p} + \epsilon_{T+1} + \hat{\theta}_1\, \epsilon_T + \ldots + \hat{\theta}_p\, \epsilon_{t-q}.$$

В этом уравнении присутствует значение ошибки из будущего ϵ_{T+1}. Неизвестно, какой будет наблюдаться шум в будущем, однако можно предполагать, что в среднем он будет равен 0. Поэтому значения будущих ошибок можно безболезненно заменить на 0. Фактически из уравнения просто удаляются все члены, которые связаны с ошибками из будущего:

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1\, y_T + \ldots + \hat{\phi}_p\, y_{T+1-p} + \hat{\theta}_1\, \epsilon_T + \ldots + \hat{\theta}_p\, \epsilon_{t-q}.$$

В уравнении также присутствуют ошибки из прошлого. Их необходимо заменить на остатки модели в этих точках, потому они являются самыми лучшими оценками ошибок из имеющихся:

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\phi}_1\, y_T + \ldots + \hat{\phi}_p\, y_{T+1-p} + \hat{\theta}_1\, \hat{\epsilon}_T + \ldots + \hat{\theta}_p\, \hat{\epsilon}_{T+1-q}.$$

Если прогноз необходимо построить не на одну точку вперёд, а, например, на две, то в формуле появляется значение ряда из будущего y_{T+1}:

$$\hat{y}_{T+2|T} = \hat{\alpha} + \hat{\phi}_1\, y_{T+1} + \ldots + \hat{\phi}_p\, y_{T+2-p} + \hat{\theta}_1\, \hat{\epsilon}_{T+1} + \ldots + \hat{\theta}_p\, \hat{\epsilon}_{T+2-q}.$$

Его необходимо заменить на прогноз $\hat{y}_{T+1|T}$.