

Лекция 9. Введение в анализ временных рядов

Введение. Данные, которые помнят свое прошлое

До сих пор мы в основном работали с перекрестными данными, где наблюдения были независимы. Но экономические данные часто представляют собой **временные ряды** - последовательность наблюдений, упорядоченных во времени (ежедневные курсы акций, ежеквартальный ВВП, месячная инфляция).

Анализ временных рядов - это совокупность методов для анализа данных, собранных в последовательные моменты времени, с целью выявления внутренней структуры и построения прогнозов.

Ключевая особенность: Наблюдения во временных рядах **зависимы**. Сегодняшняя цена акции зависит от вчерашней, а инфляция в этом месяце - от инфляции в прошлом месяце. Эта зависимость - не проблема, а источник информации!

1. Основные компоненты временного ряда

Принято считать, что временной ряд Y_t состоит из нескольких компонент:

$$Y_t = T_t + S_t + C_t + \varepsilon_t$$

где:

- **T_t (Тренд)** — долгосрочная тенденция. Может быть восходящим, нисходящим или горизонтальным (Рисунок 9.1).

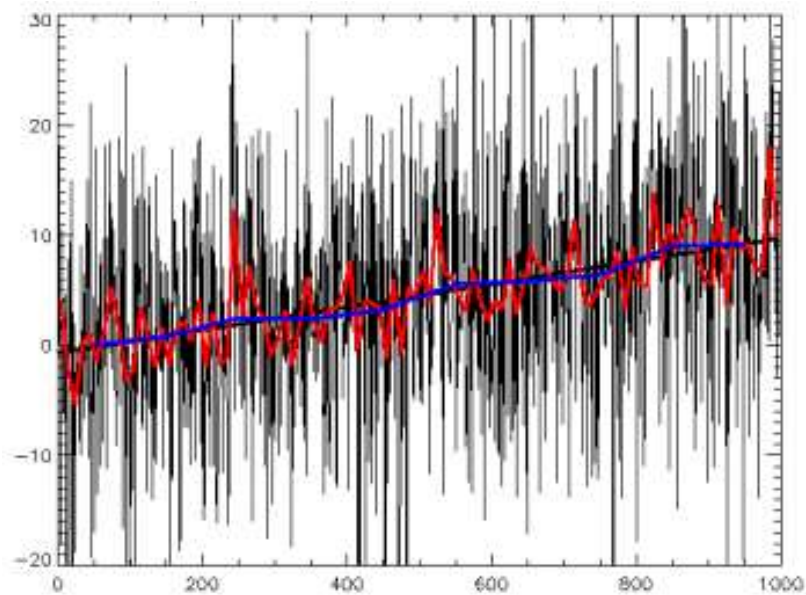


Рисунок 9.1. Тренд

- S_t (Сезонность) - периодические колебания с фиксированной частотой (год, квартал, месяц) (Рисунок 9.2).

Seasonal plot: monthly manufacture of electrical equipment

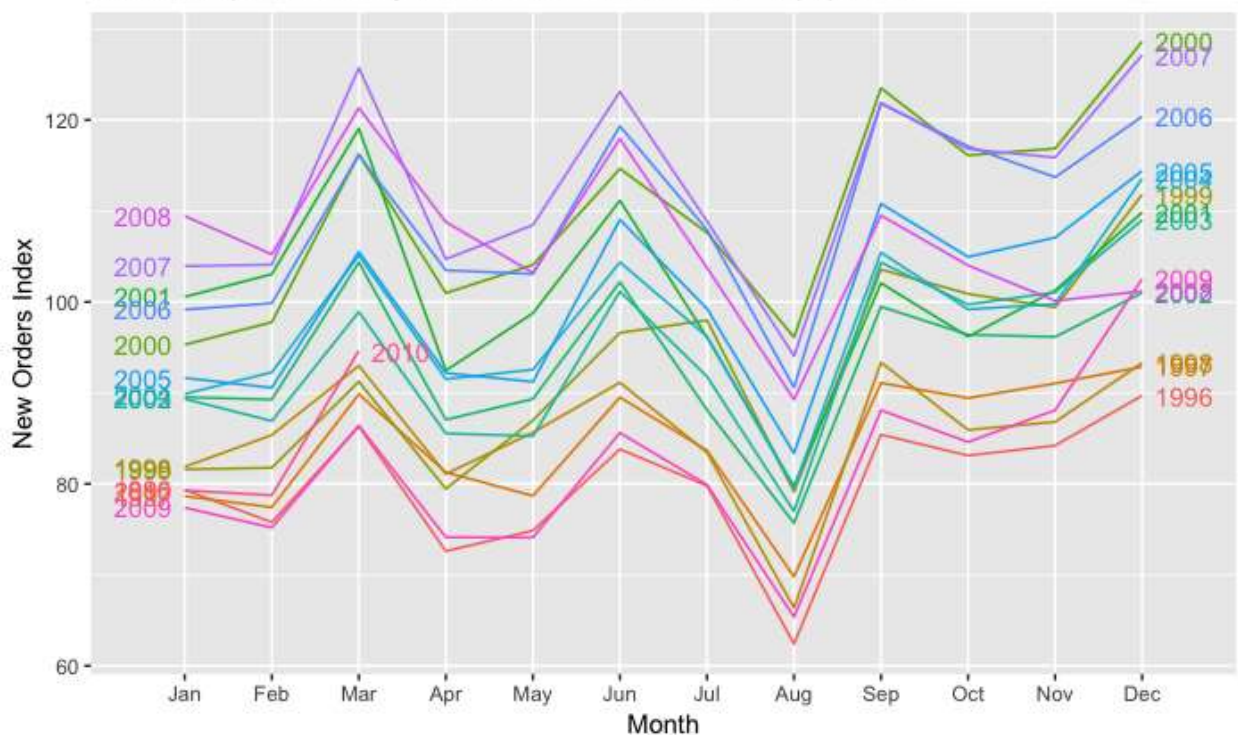


Рисунок 9.2. Сезонность

- C_t (Цикл) - долгосрочные колебания без строгой периодичности (например, экономические циклы "подъем-спад").

- ε_t (Случайная компонента) - несистематические, случайные колебания, "шум".

Подходы к анализу:

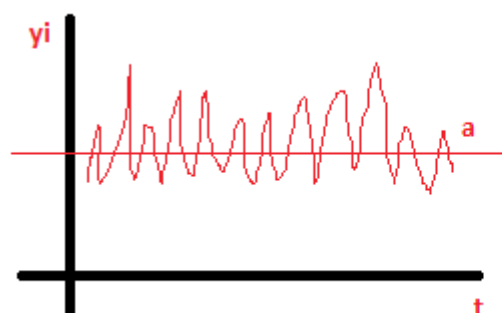
- **Моделирование компонент:** Мы можем оценить и смоделировать каждую компоненту отдельно.
- **Модели ARIMA:** Мы будем моделировать ряд как единый процесс, учитывающий его внутреннюю динамику.

2. Стационарность - краеугольный камень анализа

Стационарный временной ряд - это ряд, статистические свойства которого (среднее, дисперсия, ковариация) **не меняются с течением времени** (Рисунок 9.3).

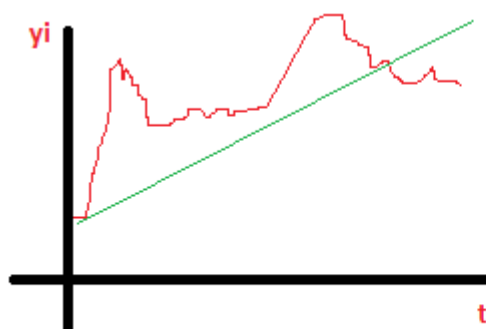
Формально ряд стационарен, если:

1. $E[Y_t] = \mu = \text{const}$ (постоянное среднее)
2. $\text{Var}[Y_t] = \sigma^2 = \text{const}$ (постоянная дисперсия)
3. $\text{Cov}[Y_t, Y_{t-k}] = \gamma_k$ (ковариация зависит только от лага k , а не от времени t)



Стационарный ряд

колебание,
отклонение от
среднего примерно
одинаково



Нестационарный ряд

часто у реального
обменного курса
такое бывает,
присутствует
блуждающий тренд -
"процесс случайного
блуждания"

Рисунок 9.3. Стационарный и нестационарный ряды

Почему стационарность так важна?

- Большинство теорий и методов разработаны для стационарных рядов.
- Если ряд нестационарен, мы можем получить **ложную регрессию (spurious regression)** - высокий R^2 и значимые коэффициенты там, где на самом деле связи нет.
- Прогнозы для нестационарных рядов ненадежны.

3. Как проверить на стационарность?

3.1. Визуальный анализ

Постройте график ряда. Если видите тренд или изменяющуюся дисперсию - ряд, скорее всего, нестационарен.

3.2. Тест Дики-Фуллера (Dickey-Fuller Test)

Самый популярный формальный тест.

- **Идея:** проверить наличие **единичного корня** — формальное условие нестационарности.
- **Модель для теста:** $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \varepsilon_t$
- **Гипотезы:**
 - $H_0: \gamma = 0$ (ряд имеет единичный корень, **нестационарен**).
 - $H_0: \gamma < 0$ (ряд **стационарен**).
- **Критерий:** сравниваем t-статистику для γ с особым критическим значением. Если t-стат. < крит. значения, отвергаем H_0 .

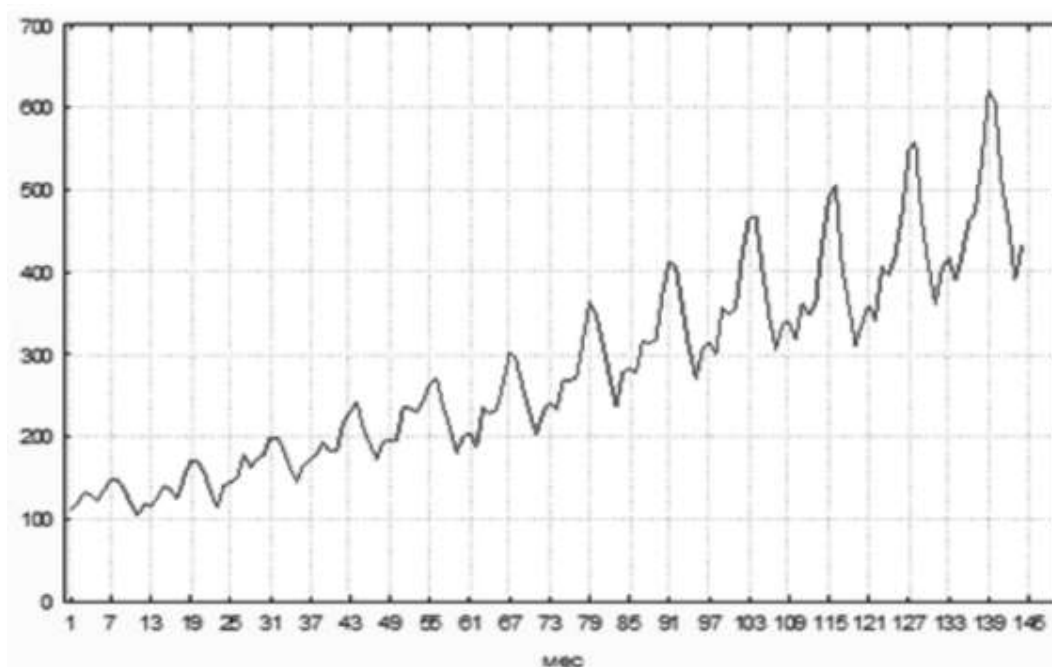
4. Как сделать ряд стационарным?

4.1. Дифференцирование (Difference)

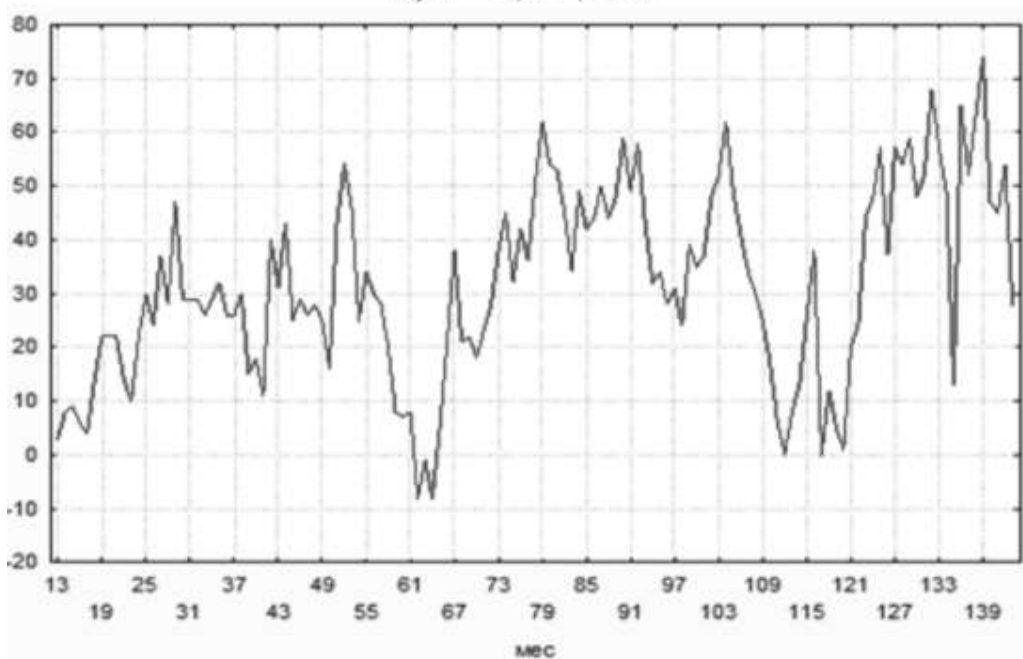
Наиболее эффективный метод для устранения тренда.

- **Первое дифференцирование:** $\Delta Y_t = Y_t - Y_{t-1}$

- Второе дифференцирование: $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$
- Обычно достаточно 1-2 дифференцирований.



а) $t = 1, \dots, 144$



б) $t = 13, \dots, 144$

Рисунок 9.4. Дифференцирование ряда

На рисунке 9.4 продемонстрировано дифференцирование ряда. Исходный ряд (сверху) имеет тренд. Продифференцированный ряд (снизу) колеблется вокруг определенного значения.

4.2. Логарифмирование

Помогает стабилизировать дисперсию (устранить гетероскедастичность), особенно если разброс данных растет со временем.

4.3. Сезонное дифференцирование

Для удаления сезонности: $\Delta_s Y_t = Y_t - Y_{t-s}$, где s - длина сезона (12 для месячных данных, 4 для квартальных).

5. Модели стационарных рядов: AR, MA, ARMA

После того как мы получили стационарный ряд, мы можем смоделировать его динамику.

5.1. Модель AR(p) — Авторегрессия

Текущее значение ряда зависит от его собственных предыдущих значений.

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

где p — порядок модели (сколько лагов использовать).

Интерпретация: Модель "инерции". Сегодняшний ВВП зависит от ВВП вчера, позавчера и т.д.

5.2. Модель MA(q) - Скользящее среднее

Текущее значение ряда зависит от прошлых случайных shocks (ошибок).

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

где q — порядок модели.

Интерпретация: Модель "реакции на шоки". Экономика реагирует на непредвиденные события (шоки) в течение нескольких периодов.

5.3. Модель ARMA(p, q)

Объединяет авторегрессию и скользящее среднее.

$$Y_t = c + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Это универсальная модель для стационарных рядов.

6. Модель ARIMA для нестационарных рядов

ARIMA(p, d, q) - это модель, которая применяется к нестационарным рядам после их преобразования в стационарные с помощью **d-кратного дифференцирования**.

- **p** — порядок авторегрессионной части (AR)
- **d** — порядок дифференцирования (I — Integrated)
- **q** — порядок скользящего среднего (MA)

Процедура построения модели ARIMA (Методология Бокса-Дженкинса):

1. **Идентификация:** построить график ряда, проверить на стационарность (тест Дики-Фуллера). Если ряд нестационарен, выбрать порядок дифференцирования **d**. Для стационарного ряда анализировать ACF и PACF графики, чтобы выбрать **p** и **q**.
2. **Оценка:** оценить параметры модели ARIMA(p,d,q) с помощью МНК или метода максимального правдоподобия.
3. **Верификация:** проверить остатки модели на "белый шум" (тест Льюнга-Бокса). Если остатки не являются белым шумом, вернуться к шагу 1.
4. **Прогнозирование:** использовать построенную модель для прогнозирования.

7. Пример: Анализ ежемесячных продаж

1. **Исходные данные:** Рост продаж с явной сезонностью.
2. **Проверка стационарности (тест Дики-Фуллера):** $p\text{-value} > 0.05 \rightarrow$ ряд нестационарен.
3. **Преобразование:**
 - Сезонное дифференцирование ($d=1, s=12$).
 - Проверяем стационарность снова: $p\text{-value} < 0.05 \rightarrow$ ряд стационарен.
4. **Идентификация модели:** Анализ ACF/PACF продифференцированного ряда указывает на $ARIMA(1,0,1)(0,1,1)\square\square$.
5. **Прогноз:** Строим прогноз на 6 месяцев вперед с доверительным интервалом.

Резюме

1. **Временные ряды** требуют специальных методов из-за зависимости наблюдений.
2. **Стационарность** - ключевое предположение. Проверяем тестом Дики-Фуллера.
3. **Дифференцирование** - основной метод достижения стационарности.
4. **ARMA-модели** описывают стационарные ряды.
5. **ARIMA-модели** обобщают подход на нестационарные ряды.
6. **Методология Бокса-Дженкинса** - системный подход к построению моделей временных рядов.
- 7.

На следующей лекции: Мы углубимся в прогнозирование с помощью моделей **ARIMA** и изучим модели с лагированной зависимой переменной (**ADL**).

Вопросы для самопроверки:

1. Почему регрессия между двумя нестационарными рядами может быть ложной?
2. В чем разница между трендом и сезонностью?
3. Если после первого дифференцирования ряд все еще нестационарен, что вы будете делать?