



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Майнор «Прикладной статистический анализ»

Временные ряды и их практическое применение (Time Series and Their Application)

Родионова Лилия Анатольевна

к.э.н., доцент департамента статистики и анализа данных НИУ ВШЭ

LRodionova@hse.ru



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

*Родионова Лилия
Анатольевна*

к.э.н., доцент

*Департамент статистики и
анализа данных
Факультет экономических наук
НИУ ВШЭ (Москва)*

LRodionova@hse.ru

Web: <https://www.hse.ru/staff/LRodionova>





Организация курса

- **3-4 Модуль 2020-2021 уч.г.г.**
- **Лекции 28 часов (14 лекций)**
- **Практика 48 часов (24 занятия)**
- **Контроль: СР1, СР2, текущие ДЗ, текущие тесты.**
- **Стат пакеты: Stata, Gretl (на занятиях), дома – любой пакет**
- **Материалы лекций, задания и отчеты в LMS**
- **Работа с учебными ассистентами:**
 - Артемова Полина
 - Васильева Наталья
 - Юдина Валерия
 - Никонова Екатерина



Оценка по курсу

Промежуточная аттестация

% Промежуточная аттестация (4 модуль)

0.200 активность на семинарах

0.200 Самостоятельная работа 1

0.200 Самостоятельная работа 2

0.200 Текущие домашние работы

0.200 Текущие тесты

Условие пересдачи: СР1, СР2



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Рекомендуемая литература:



- Айвазян С.А. Методы эконометрики: учебник— М.: Магистр: ИНФРА-М, 2010.



- Анализ временных рядов (курс лекций) / Канторович Г. Г.// Экономический журнал ВШЭ, 2002-2003.

<http://www.hse.ru/mag/economics>

[2002] Т. 6 № 1. С. 85–116

[2002] Т. 6 № 2. С. 251–273

[2002] Т. 6 № 3. С. 379–401

[2002] Т. 6 № 4. С. 498–523

[2003] Т. 7 № 1. С. 79–103

Рекомендуемая литература:



- Подкорытова, О. А. Анализ временных рядов : учебное пособие для бакалавриата и магистратуры. — М. : Издательство Юрайт, 2016. — 266 с. — (Серия : Бакалавр и магистр. Модуль.).

Рекомендуемая литература:



- Носко В.П. Эконометрика Книга 1. — М.: Изд. дом «Дело», 2011.



- Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс.- М.: Дело, 2004. – 576 с.

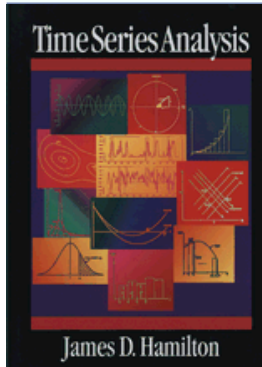
Электронные ресурсы:

<http://www.twirpx.com/files/financial/apvr>

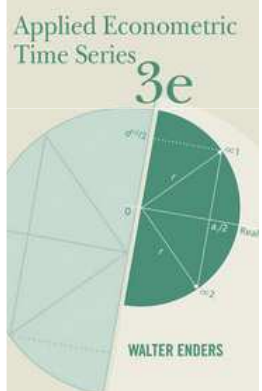


НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

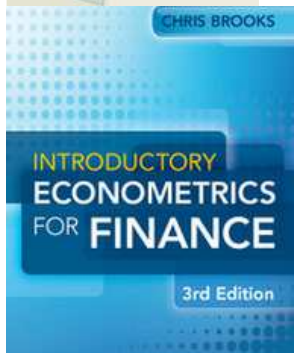
Рекомендуемая литература:



Hamilton James D. Time Series Analysis. Princeton University Press, 1994.



Enders W. Applied Econometric Times Series. Wiley, 2009.

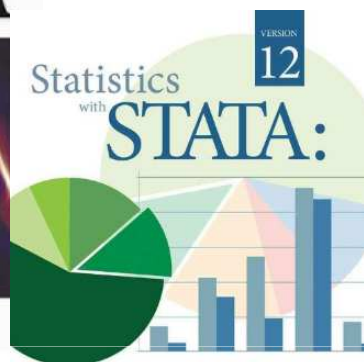


Brooks C. Introductory Econometrics for Finance, 2014. (VAR, GARCH, switching models)



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Рекомендуемая литература (стат. пакеты):



Lawrence C. Hamilton

- Hamilton L.C. Statistics with STATA. 2004, 2012.

<http://www.twirpx.com/file/825870/>

- Куфель Т. Эконометрика: решение задач с применением пакета программ GRETL, 2007.

<http://www.twirpx.com/file/970212/>

Сайт программы (Free Software):
<http://gretl.sourceforge.net/>





Обозначения. Греческий алфавит

Прописные	Строчные	Название
A	α	альфа
B	β	бета
Γ	γ	гамма
Δ	δ	дельта
E	ε	эпсилон
Z	ζ	дзета
H	η	эта
Θ	θ, ϑ	тета
I	ι	йота
K	κ, ϰ	каппа
Λ	λ	лямбда
M	μ	мю
N	ν	ню
Ξ	ξ	кси
O	ο	омикрон
Π	π	пи
P	ρ	ро
Σ	σ, ς	сигма
T	τ	тау
Υ	υ	ипсилон
Φ	φ, ϕ	фи
X	χ	хи
Ψ	ψ	пси
Ω	ω	омега



Введение



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

История прогнозирования

Цель анализа временных рядов – прогнозирование

История:

-Еврейский пророк Исаия написал примерно в 700 г. до н.э. «Tell us what the future holds, so we may know that you are gods».(Isaiah 41:23)

(«Расскажите нам, что нас ждет в будущем, чтобы мы знали, что вы боги»)

-В древнем Вавилоне, синоптики предсказывали будущее, основываясь на распределении личинок в печени гнилой овцы.

-К 300 г. до н.э. люди, желающие получить прогнозы, отправлялись в Дельфы в Греции, чтобы проконсультироваться с Оракулом, который предоставлял свои прогнозы, находясь в состоянии алкогольного опьянения.

<https://otexts.com/fpp2/intro.html>



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

История прогнозирования

- император Константин, указ в 357 г. н.э., запрещающий кому-либо «консультироваться с предсказателем, математиком... Пусть любопытство предсказывать будущее будет навсегда заглушено»
- Англия, 1736 г., запрет на прогнозирование, преступление - взимать деньги за предсказания. Наказание - три месяца каторжных работ!
- ! В нашем курсе рассмотрим современные **статистические** подходы прогнозирования.



Что можно прогнозировать?

- Прогнозирование** - важная помощь для эффективного и результативного планирования.
- Прогноз спроса на электроэнергию:** строить ли еще одну электростанцию в ближайшие пять лет;
- Прогноз объема звонков:** составление графика работы персонала в колл-центре на следующей неделе;
- Прогноз спроса на товар:** запас товара на складе.
- Типы прогнозов:** Прогнозы могут потребоваться на несколько лет (для капиталовложений) или на несколько минут (для телекоммуникационной маршрутизации).



Условия прогнозирования

Предсказуемость события зависит от :

- Факторов, влияющих на событие;
- Доступности данных;

Что можно прогнозировать?

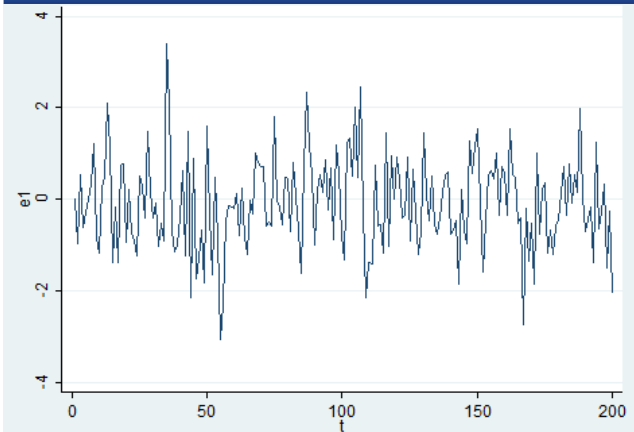
Пример 1. Прогнозы спроса на электроэнергию

Факторы: спрос на электроэнергию зависит от температуры, времени года, суток, в меньшей степени влияет на изменения календаря (праздники и экономические условия). **Доступность данных** - спрос на электроэнергию и погодные условия, прогноз может быть точным.

Пример 2. Прогнозирование курсов валют. Выполняется только *одно* из условий: имеется множество доступных данных. Ограничение понимания факторов, влияющих на обменные курсы, и прогнозы обменного курса напрямую влияют на сами курсы.



Структура курса: какие модели будем изучать?

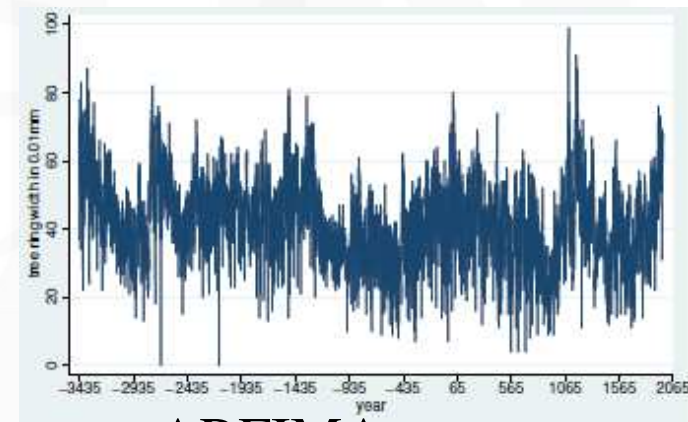
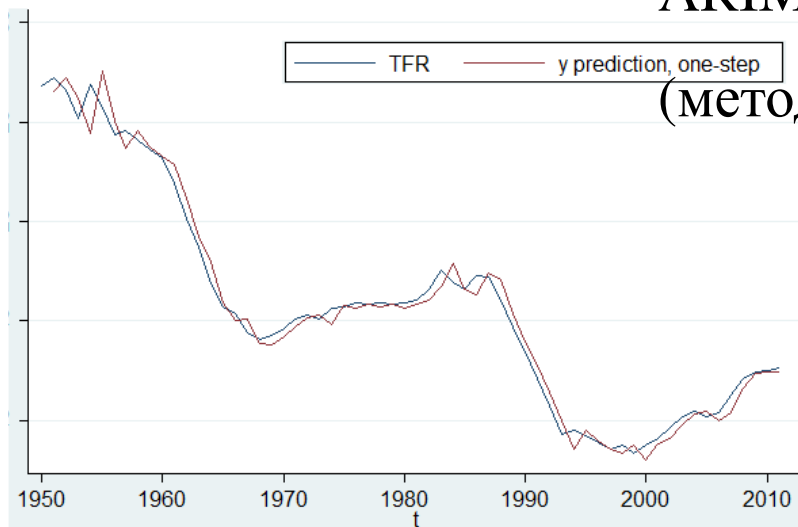


Модели стационарных ВР

ARMA-модели

ARIMA-модели

(методология Бокса-Дженкинса) 70-е г.г.



ARFIMA-модели



Класс ARIMA-моделей: примеры использования

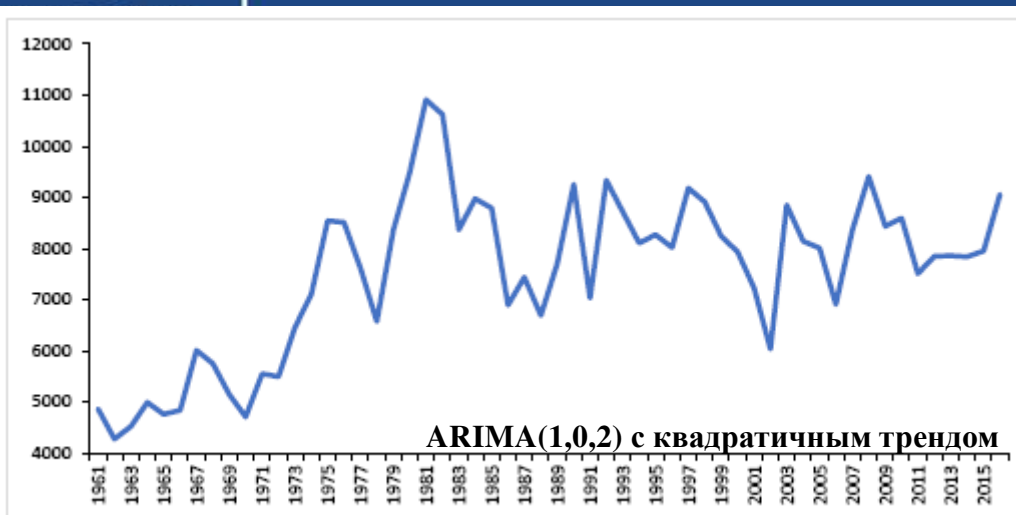
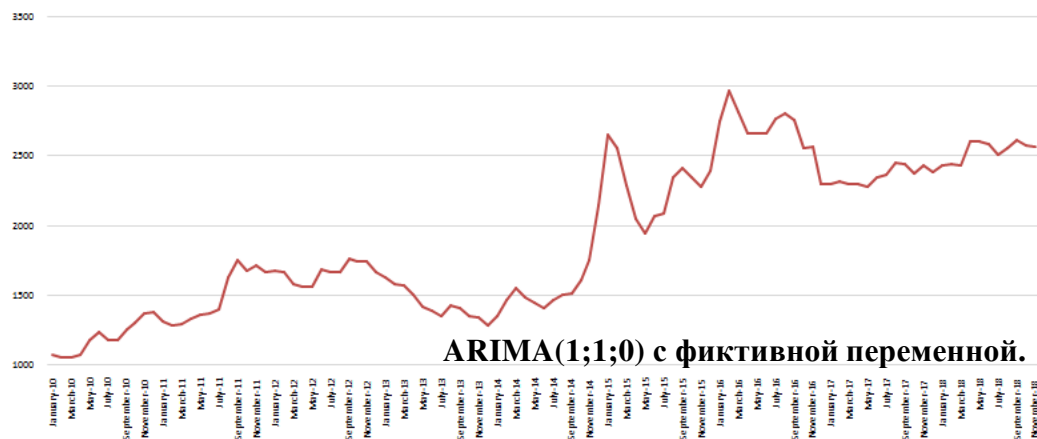
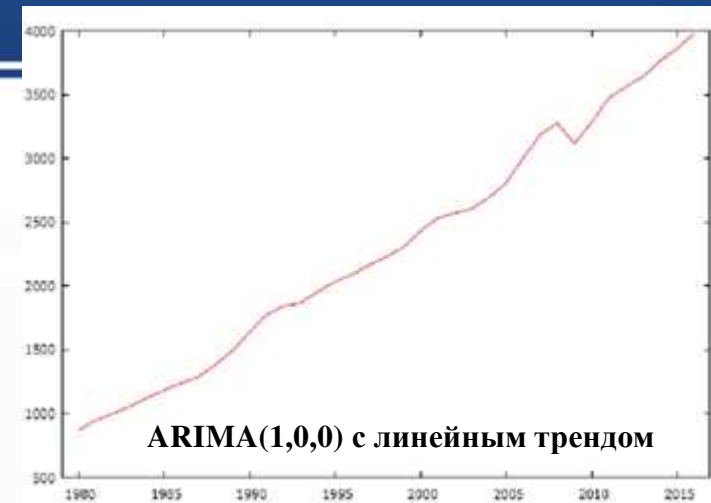


Рис. 1. Чистая стоимость производства пшеницы в США
(в млн. постоянных долл. США 2004 - 2006) по годам 1961 - 2016
Источник: данные FAOSTAT (<http://www.fao.org/faostat>)



Динамика цен на золото (руб./грамм) 2010 -2018



Значения реального ВВП Испании (млрд. евро)

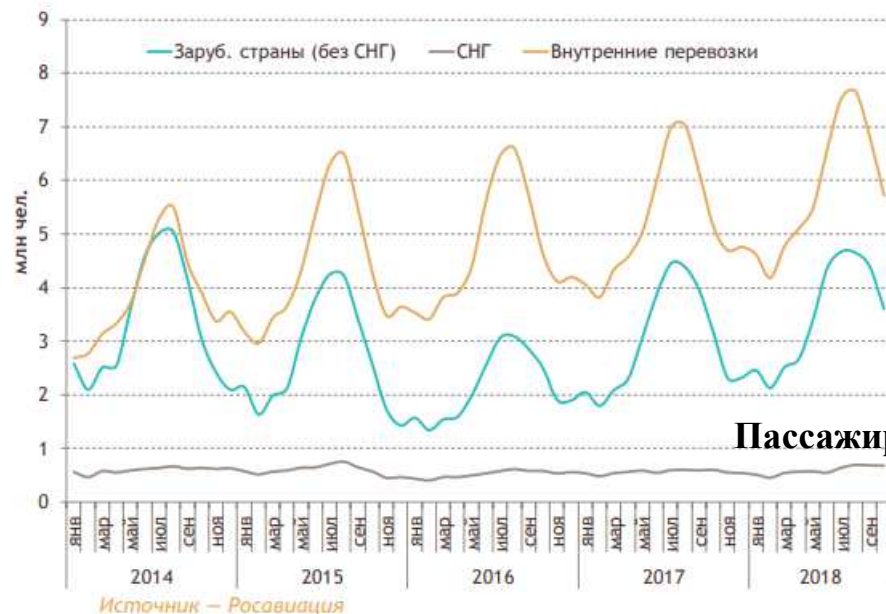




НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Структура курса: какие модели будем изучать?

Динамика потребления электроэнергии в ЕЭС
России по мес 2010–2012 г.г.



Анализ и моделирование

сезонности:

-SARIMA-модели

-трендсезонные

-модели с фиктивными
переменными,

гармоническими составляющими

-адаптивные модели

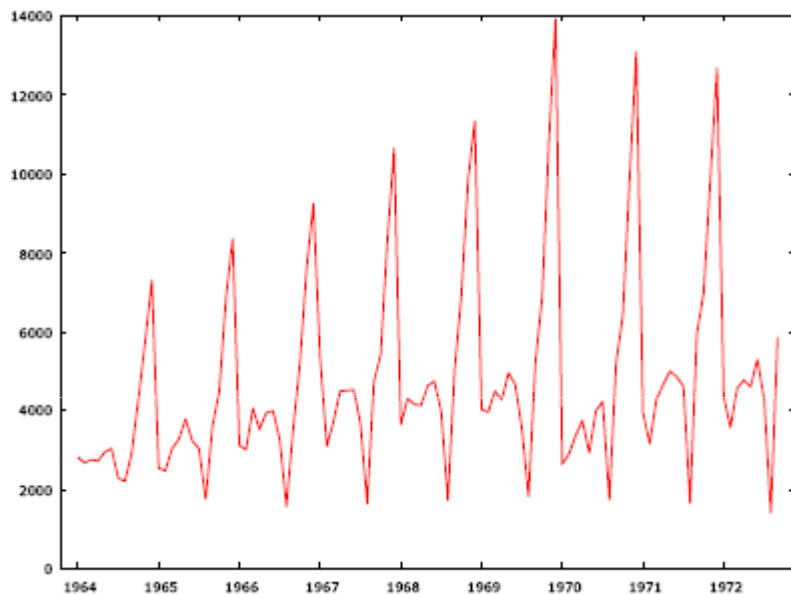
Пассажирские авиаперевозки в России, млн чел.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

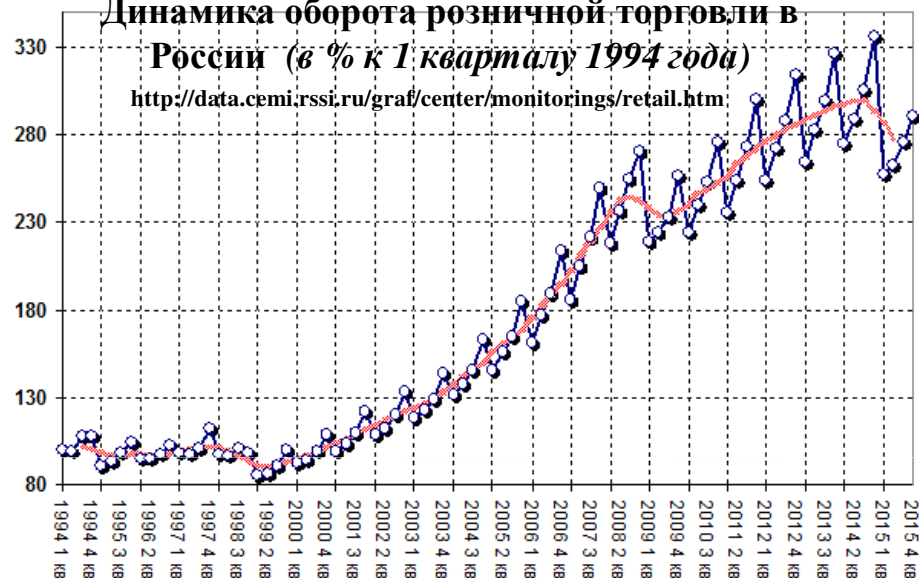
Класс SARIMA-моделей: примеры использования

Ежемесячные продажи шампанского Perrin Freres
(млн. долл США)

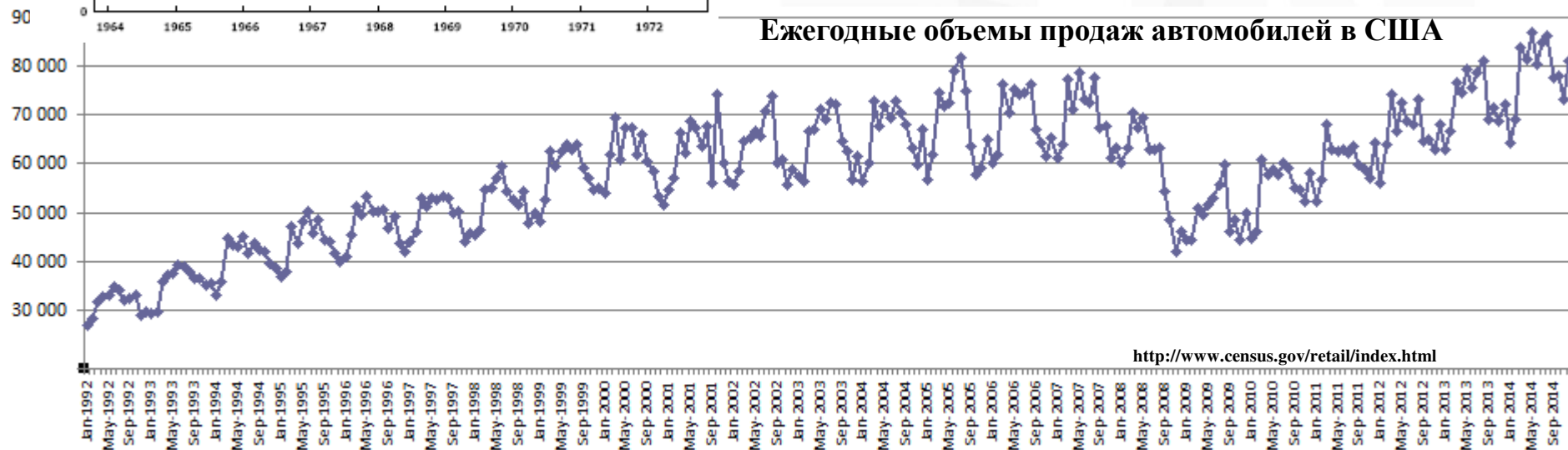


Динамика оборота розничной торговли в
России (в % к 1 кварталу 1994 года)

<http://data.cemi.rssi.ru/graf/center/monitorings/retail.htm>



Ежегодные объемы продаж автомобилей в США



<http://www.census.gov/retail/index.html>

Структура курса: какие модели будем изучать?

Кластеризация волатильности

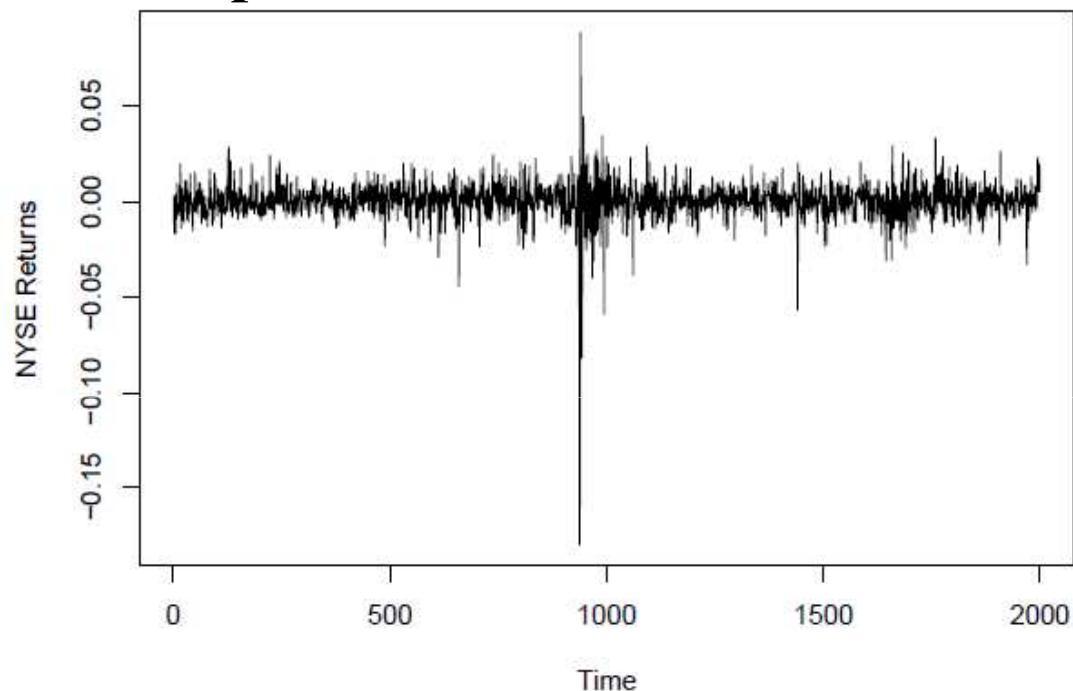


Fig. 1.4. Returns of the NYSE. The data are daily value weighted market returns from February 2, 1984 to December 31, 1991 (2000 trading days). The crash of October 19, 1987 occurs at $t = 938$.

- 19 октября 1987. Чёрный понедельник: индекс Доу-Джонса пережил самое большое падение в истории — на 22,6 %.

ARCH and GARCH models (Engle, 1982; Bollerslev, 1986)

Нью-Йоркская фондовая биржа

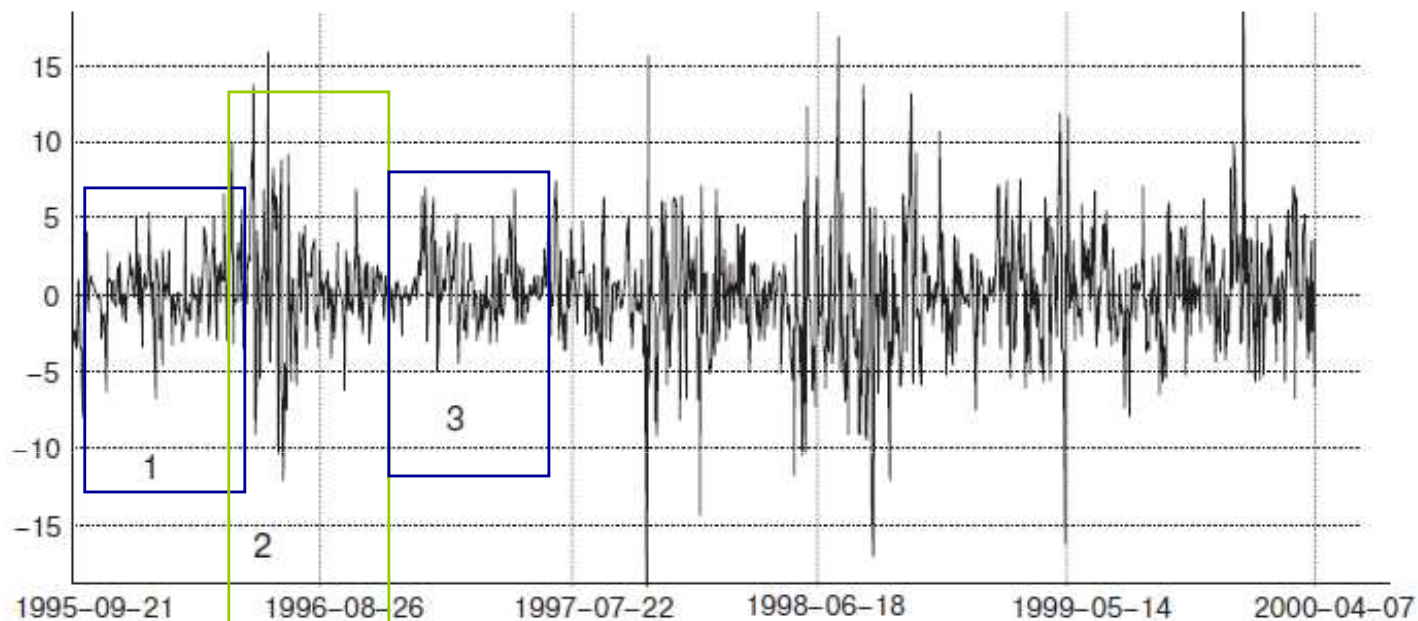
New York Stock Exchange,





НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Структура курса: какие модели будем изучать?



Темпы прироста индекса РТС с 21 сентября 1995 г. по 7 апреля 2000 г., в процентах.

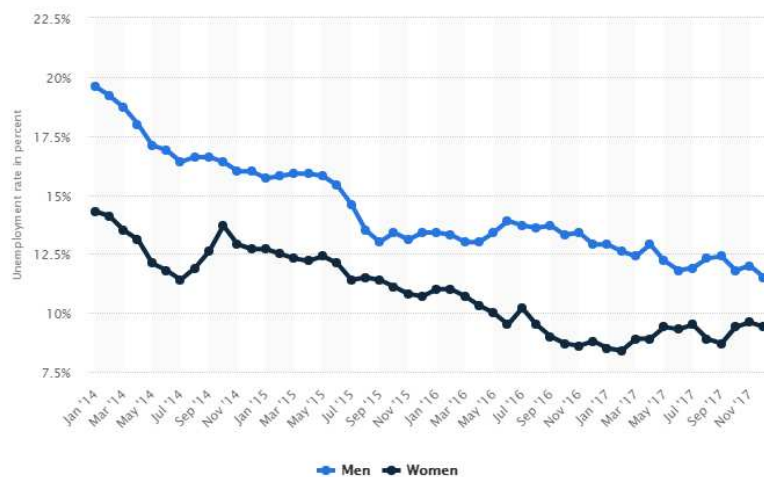
Модели с авторегрессионной условной гетероскедастичностью
ARCH (AutoRegressive Conditional Heteroscedasticity),



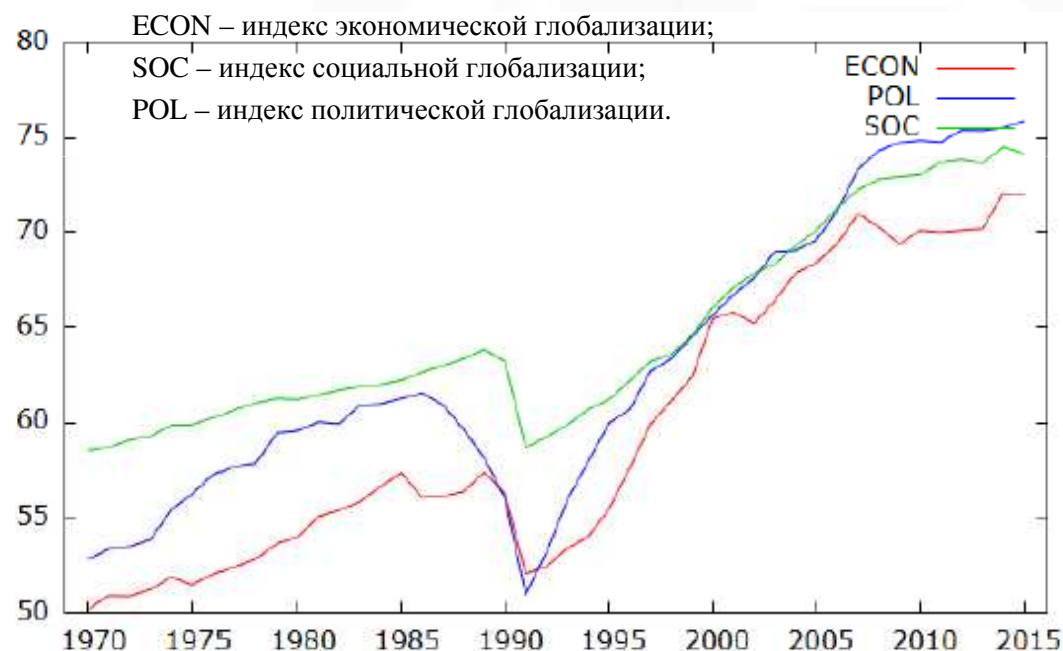
Структура курса: какие модели будем изучать?



Модели многомерных временных рядов:
VAR-модели
Модели коинтеграции (ECM)



Monthly youth unemployment rate (18-24 years old) in UK, by gender



Динамика индексов глобализации

Тема 1. Введения в анализ одномерных временных рядов

Временной ряд: основные понятия, определения, характеристики. Простейшие примеры стационарных и нестационарных временных рядов (белый шум, временной ряд с линейным трендом, случайное блуждание, случайным блужданием с дрейфом) и их характеристики. Основные составляющие временного ряда.

Исследование и моделирование неслучайной составляющей временного ряда: основные типы трендов (детерминированный и стохастический); проверка наличия тренда во временных рядах; методы выделения тренда.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Сложность курса

- От простого к сложному
- Необходимость чтения документации к стат пакетам, статей по анализу ВР в журналах, учебников по анализу ВР

Пример

Impulse Response Function

An impulse response function (IRF) of a time series model (or dynamic response of the system) measures the change in a variable when a variable is shocked by an impulse. In other words, the IRF at time t is the derivative of the responses at time t when an innovation was shocked, $t \geq t_0$.

Consider a `numseries-D VEC(p-1)` model for the multivariate response variable y_t . In lag operator notation, the

$$\Gamma(L)y_t = c + dt + \beta x_t + \varepsilon_t,$$

where $\Gamma(L) = I - \Gamma_1 L - \Gamma_2 L^2 - \dots - \Gamma_p L^p$ and I is the `numseries-by-numseries` identity matrix.

In lag operator notation, the infinite lag MA representation of y_t is:

$$\begin{aligned} y_t &= \Gamma^{-1}(L)(c + \beta x_t + dt) + \Gamma^{-1}(L)\varepsilon_t \\ &= \Omega(L)(c + \beta x_t + dt) + \Omega(L)\varepsilon_t. \end{aligned}$$

The general form of the IRF of y_t shocked by an impulse to variable j by one standard deviation of its innovation is

$$\psi_j(m) = C_m e_j.$$

https://www.mathworks.com/help/econ/vecm.irf.html#mw_9ad2b0fa-23d2-44b5-8e39-741684571ef8



Тема 1.

Введение в анализ одномерных временных рядов



Структура данных

**Пространственная
выборка**

$$Y_i$$

Временной ряд

$$Y_t$$

Определение 1.1.

Временной ряд (time series) – совокупность наблюдений экономической величины в различные моменты времени.

(Канторович)

Другие названия в литературе: *ряд динамики, динамический ряд.*

- Элементы ВР: время (t) и значение показателя (y_t)
- Классификация ВР: по форме представления уровней; по характеру временного параметра (из курса статистики)
- Случайный процесс и ВР: случайный процесс – случайные функции от времени $\{Y_t, t = \overline{1, T}\}$
- ВР – это реализация случайных процессов (или говорят, ВР порождается стохастическим процессом).

$$\dots, Y_{-1}, Y_0, \underline{Y_1, Y_2, \dots, Y_T}, Y_{T+1}, Y_{T+2}, \dots$$

Наблюдаемая выборка



Временной ряд: основные понятия и определения

Определение 1.2.

Временной ряд (time series) – ряд наблюдений $Y_{t_1}, Y_{t_2}, \dots, Y_{t_N}$, анализируемой сл. величины $\xi(t)$, произведенных в последовательные моменты времени t_1, t_2, \dots, t_N .

(Айвазян)

Проблема .

Если основные статистические характеристики случ. процесса со временем меняются, то по поведению ВР (как реализации) ничего сказать нельзя

→ необходимо рассмотрение узкого класса случайных процессов (стационарного)

Далее будем рассматривать: **одномерные, дискретные с равноотстоящими моментами наблюдений, случайные ВР.** $Y_t, t \in [t_o, T], \Delta = t_i - t_{i-1} = \text{const}$
характеристика периода, **лаг**

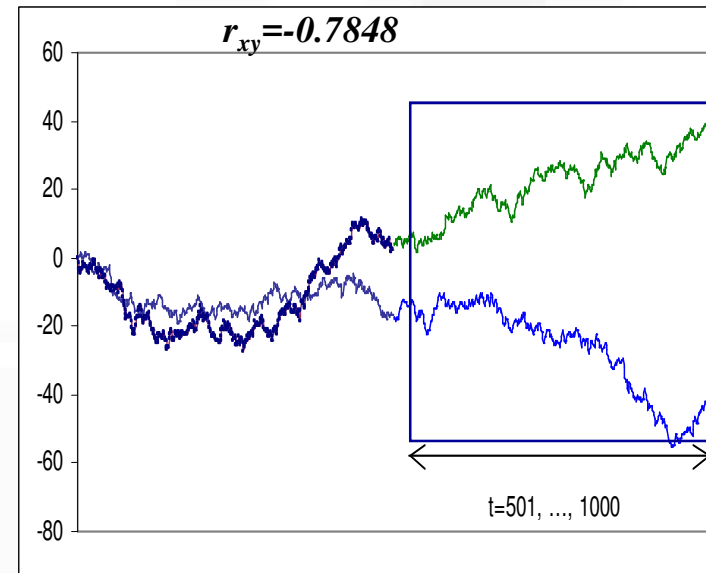
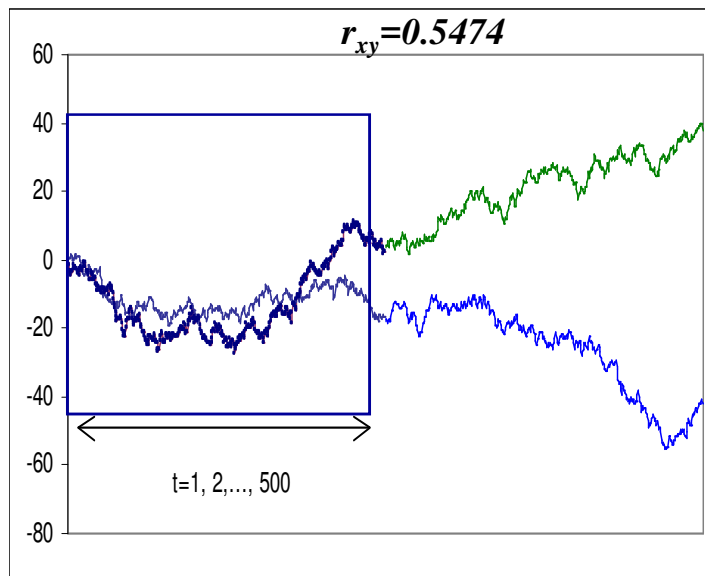


Лag=1 год



Лag=1 день

- **Различия.** 1. Члены ВР статистически зависимы.
2. Члены ВР не являются одинаково распределенными.
- **Проблема ложной корреляции:** $x_t = x_{t-1} + \varepsilon_t$, $y_t = y_{t-1} + v_t$.



! Нельзя рассматривать свойства и правила статистического анализа случайной выборки на ВР

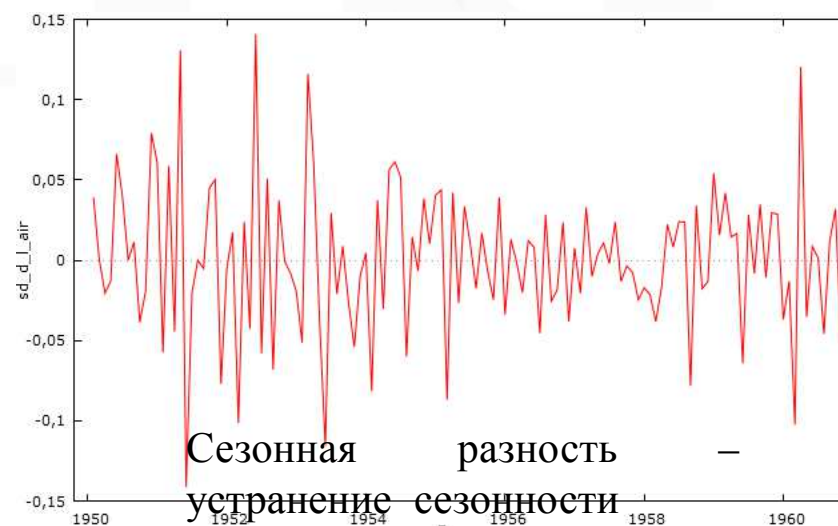
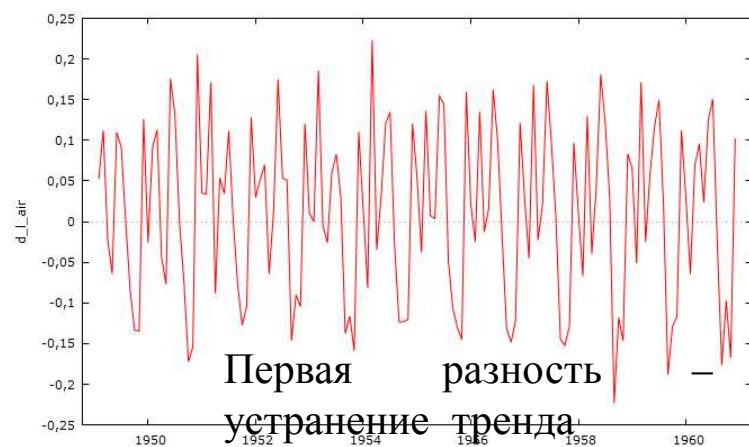
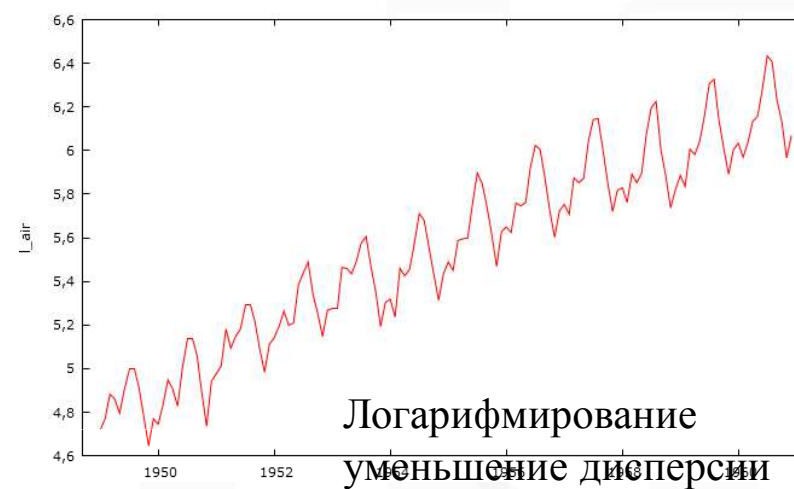
Зачем преобразовывать ряд?

- Выделение интересующих временных компонент.
- Удаление действия мешающих компонент (например, сезонности).
 - *Сезонная корректировка*
- Приведение ряда к стационарному виду (например, подход Бокса-Дженкинса).
 - *дифференцирование (взятие разности), детрендирование.*
- Уменьшение ложных эффектов регрессии.
 - *дифференцирование* в случае ложных эффектов при коинтеграции
- Стабилизирование изменчивости, которая растет с уровнем ряда.
 - *логарифмирование*
- Необходимость сопоставимости рядов.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ТРАНСПОРТА

ВР: преобразования, пример



Первое запаздывание (первый лаг) ВР Y_t – это Y_{t-1} .

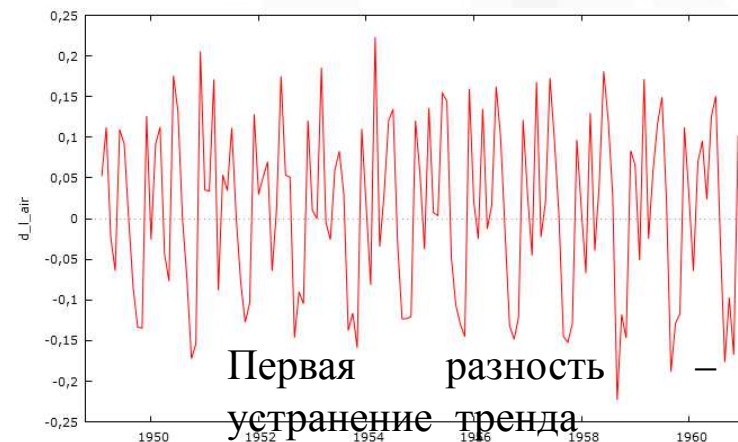
j-е запаздывание (j-й лаг) ВР Y_t – это Y_{t-j} .

Первая разность ВР $\Delta Y_t = Y_t - Y_{t-1}$ – это его изменение (приращение) между периодами (t-1) и t.

(Позже введем *лаговый и разностный* операторы)

Вторая разность

$$\Delta^2 y_t = \Delta \Delta y_t = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$



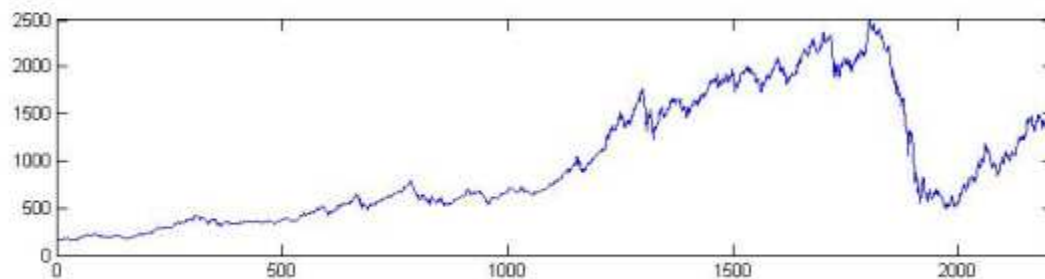
ВР: преобразования

Первая разность логарифма (логарифмическая разность) $\Delta \ln Y_t = \ln Y_t - \ln Y_{t-1}$ – это процентное изменение Y_t между периодами $(t-1)$ и t , равно $100 \cdot \Delta \ln Y_t$.

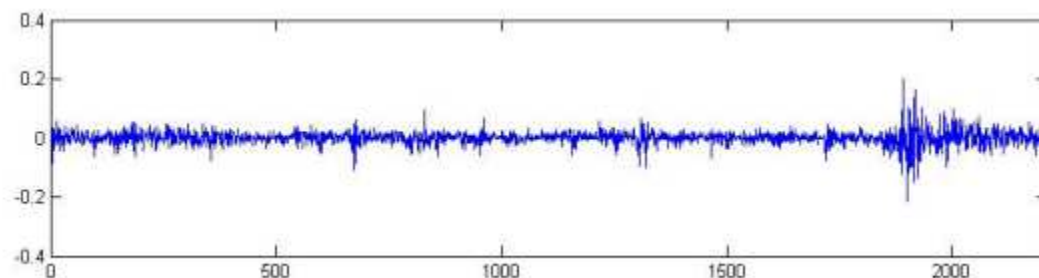
аналог темпа прироста:

$$T_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \cdot 100\%$$

Скорость изменения ценового ряда называется *доходностью*. В *финансовой эконометрике* вместо ценовых рядов обычно используются ряды доходностей.



Дневные цены закрытия индекса РТС



логарифмическая доходность

доходность актива

$$T_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \cdot 100\%$$

Логарифмическая разность (Первая разность логарифма)

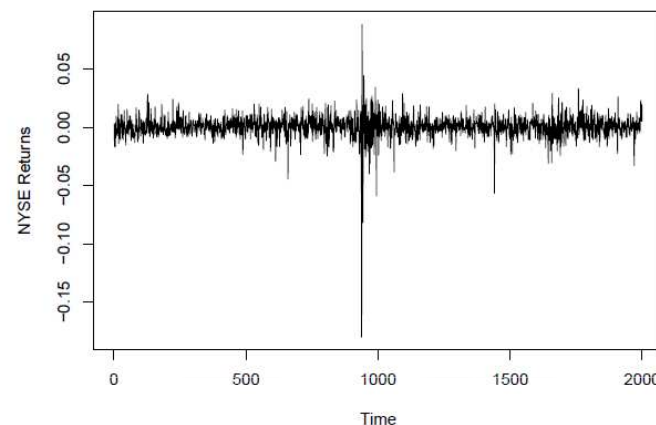
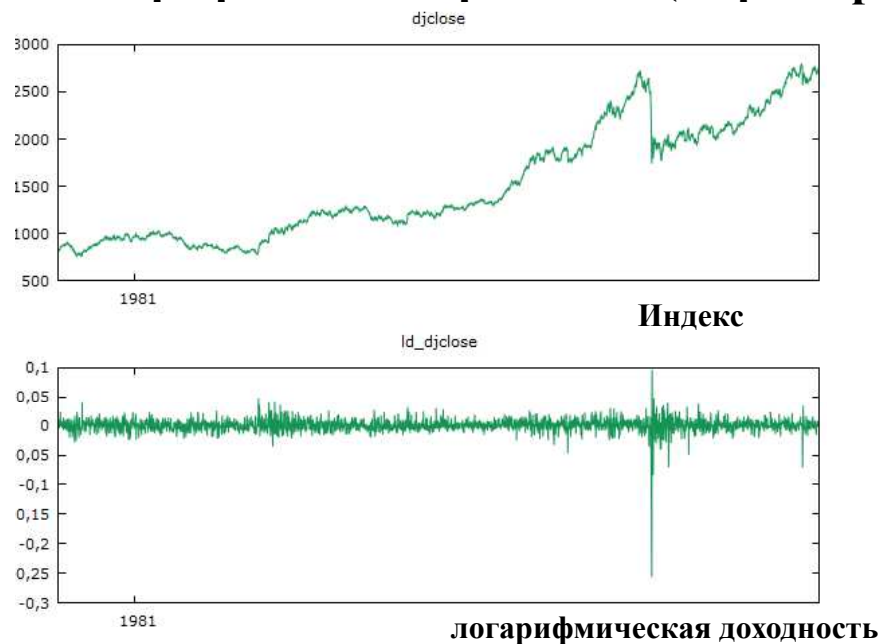


Fig. 1.4. Returns of the NYSE. The data are daily value weighted market returns from February 2, 1984 to December 31, 1991 (2000 trading days). The crash of October 19, 1987 occurs at $t = 938$.

Нью-Йоркская фондовая биржа
New York Stock Exchange,



•19 октября 1987. Чёрный понедельник: индекс Доу-Джонса пережил самое большое падение в истории — на 22,6 %.

Индекс Доу-Джонса - старейший из существующих американских рыночных индексов (с 1884 г.), был создан для отслеживания развития промышленной составляющей американских фондовых рынков, индекс охватывает 30 крупнейших компаний США, рассчитывается как масштабируемое среднее цен на акции.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ВР: показатели динамики ряда

Численность
России населения

Пример.

Годы	Все население, млн. человек	в том числе	
		городское	сельское
2007	142,8	104,7	38,1
2008	142,8	104,9	37,9
2009	142,7	104,9	37,8
2010	142,9	105,3	37,6
2011	142,9	105,4	37,5
2012	143,0	105,7	37,3
2013	143,3	106,1	37,2
2014	143,7	106,6	37,1
2015	146,3	108,3	38,0
2016	146,5	108,6	37,9
2017	146,8	109,0	37,8
2018	146,9	109,3	37,6

Первая разность $\Delta Y_t = Y_t - Y_{t-1}$ – это изменение (приращение) между периодами (t-1) и t.

Первая разность логарифма $\Delta \ln Y_t = \ln Y_t - \ln Y_{t-1}$ – это процентное изменение Y_t между периодами (t-1) и t. Равно $100 \cdot \Delta \ln Y_t$.

Темпа прироста: $T_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \cdot 100\%$

Как изменилась численность населения в 2018г.?

$$\Delta Y_{t=2018} = Y_{2018} - Y_{2017} = 146,9 - 146,8 = 0,1$$

$$T_{t=2018} = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \cdot 100\% = \frac{146,9 - 146,8}{146,8} \cdot 100\% = 0,07\%$$

http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/demography/#

4 типа факторов (Айвазян):

(А) Долговременные → общая тенденция Y_t

(Б) Сезонные → периодически повторяющиеся в определенное время года колебания Y_t

(В) Циклические → изменения Y_t , обусловленные действием долговременных циклов

(Г) Случайные (не поддаются учету и регистрации)

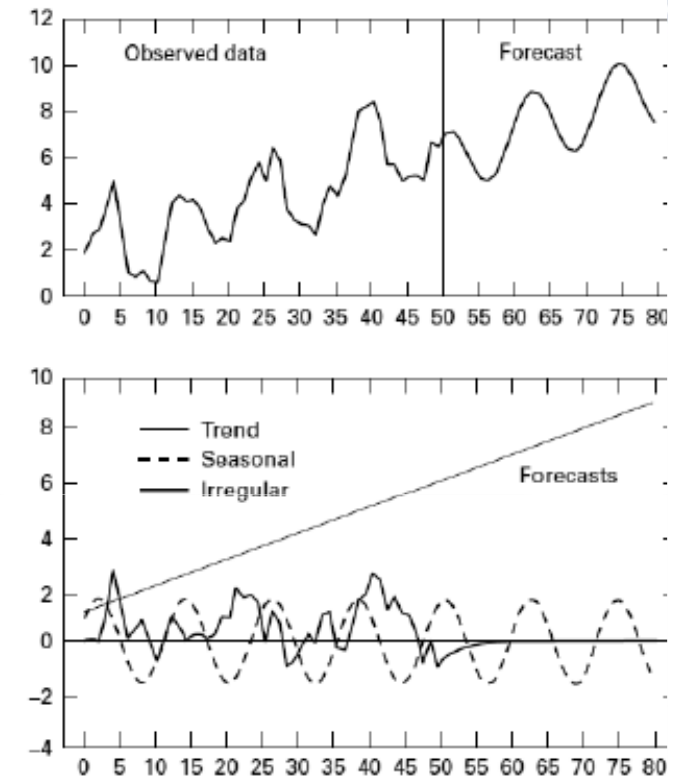


FIGURE 1.1 Hypothetical Time Series

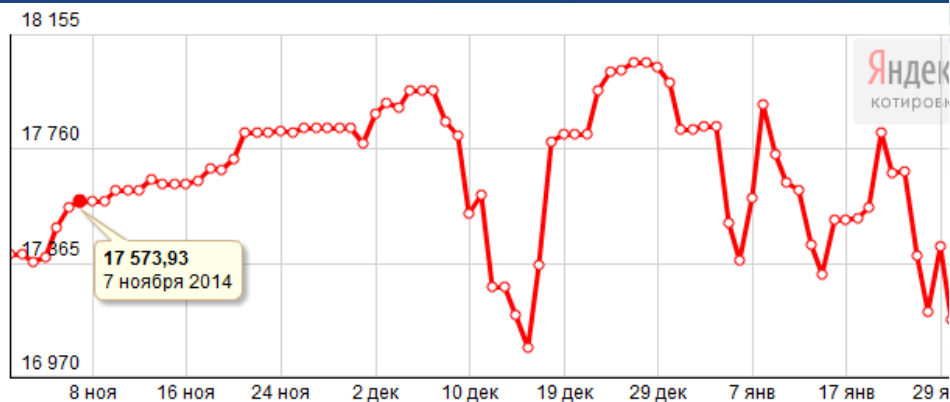
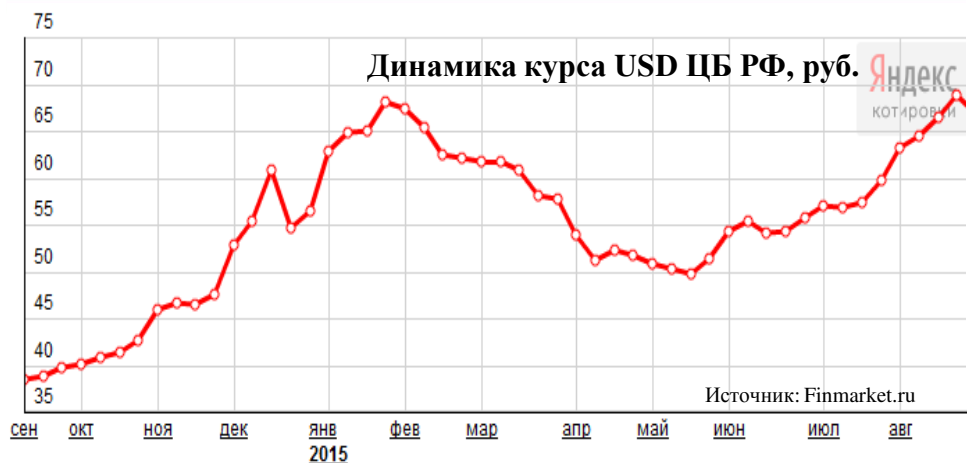
Enders

$$Y_t = \underbrace{\chi(A)f_{mp}(t)}_{\text{Трендовая компонента}} + \underbrace{\chi(B)\varphi(t)}_{\text{Сезонная компонента}} + \underbrace{\chi(B)\psi(t)}_{\text{Циклическая компонента}} + \underbrace{\varepsilon_t}_{\text{Случайная компонента}}$$



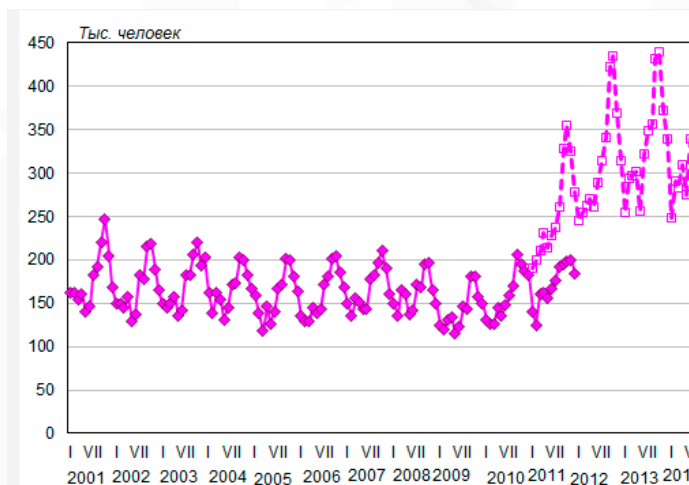
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Компоненты ВР



Динамика индекса Dow (DJIA) Доу Джонса

<http://news.yandex.ru/quotes/12.html>

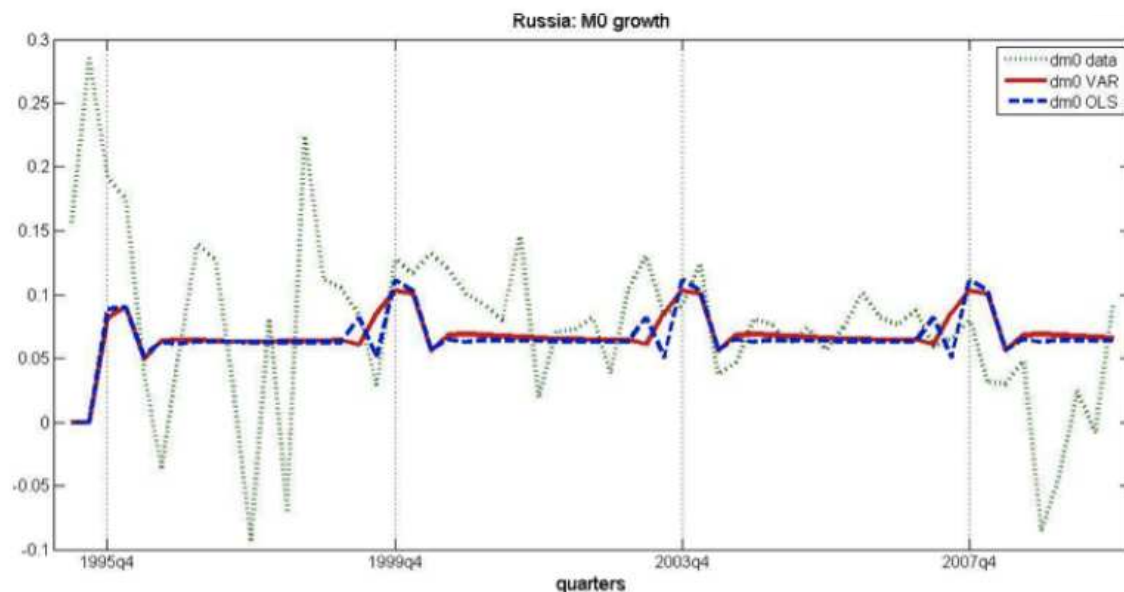


Число мигрантов, перемещающихся в пределах России

Компоненты ВР

Рост денежной массы в России в зависимости от выборов: манипулируют ли политики монетарными инструментами для того, чтобы выиграть выборы в новых демократиях?

Источник: Журнал Квантиль №11. Бурковская Анастасия. Монетарные политические бизнес-циклы: новые демократии <http://quantile.ru/11/N11.htm>



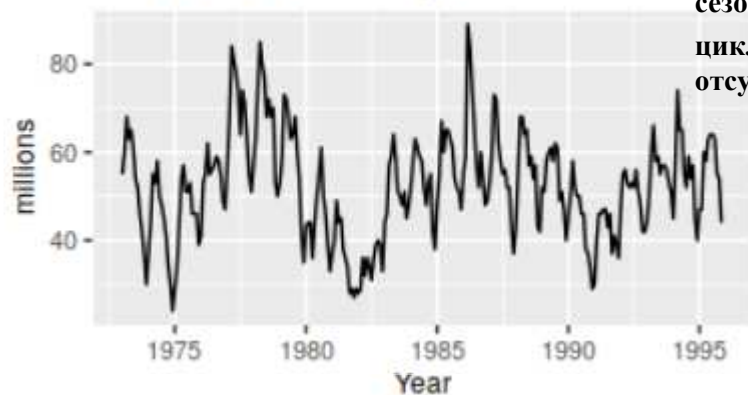
Денежная масса — совокупность наличных денег, находящихся в обращении, и безналичных средств на счетах, которыми располагают физические и юридические лица и государство.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ЦЕНТР ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

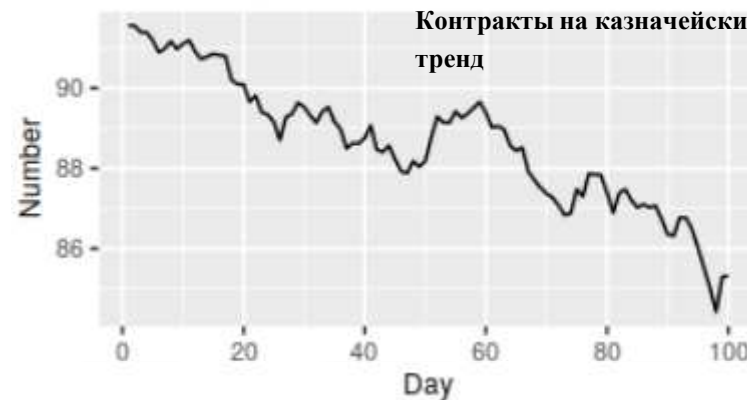
Генезис наблюдений, образующих ВР

Sales of new one-family houses, USA



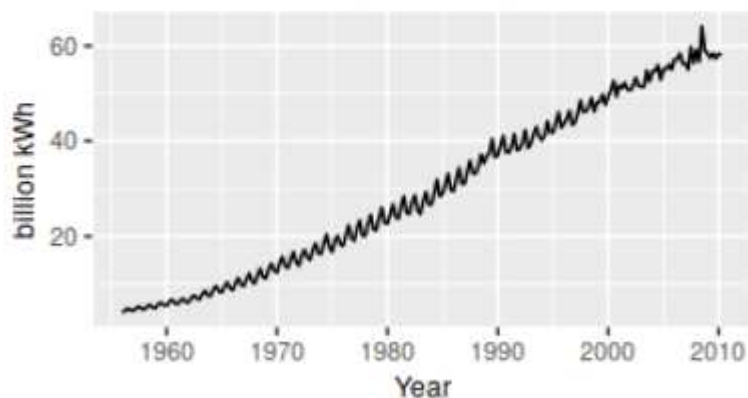
сезонность,
цикл с периодом 6–10 лет,
отсутствие тренда

US treasury bill contracts

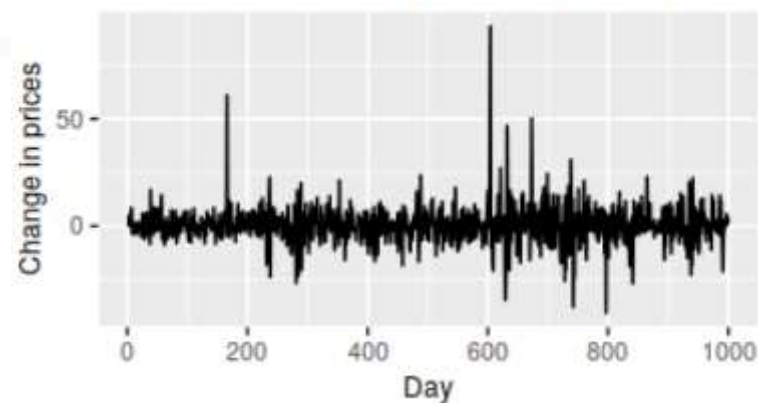


Контракты на казначейские вексел
тренд

Australian quarterly electricity production



Google daily changes in closing stock price



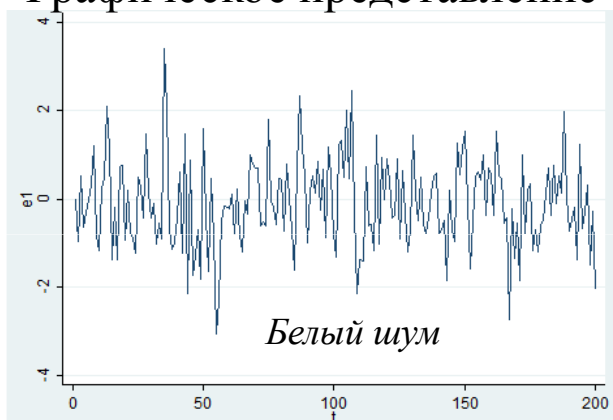


ВР и их характеристики: среднее

Теоретические

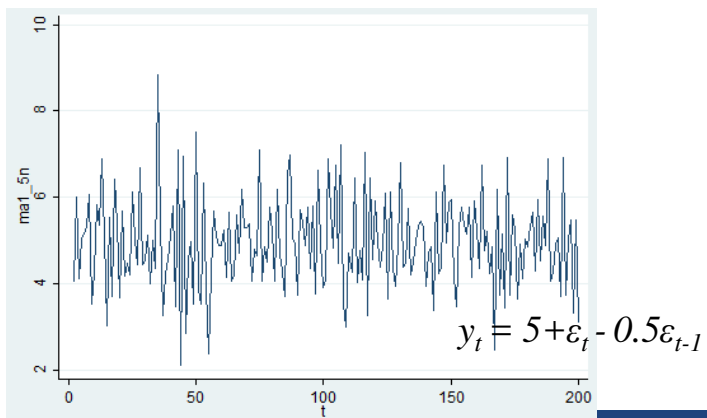
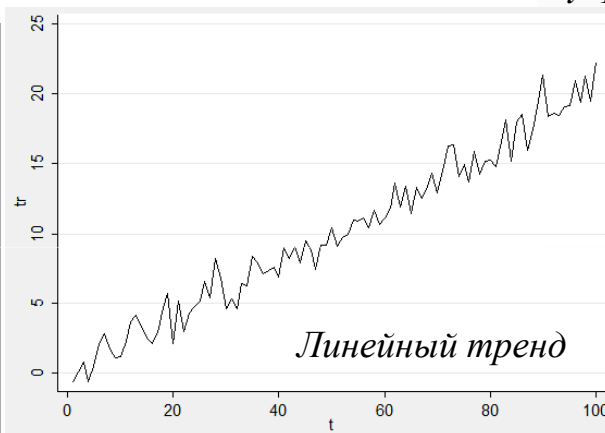
$$E(Y_t) = \mu$$

Графическое представление



Выборочные

$$\hat{\mu} = \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$





ВР и их характеристики: дисперсия

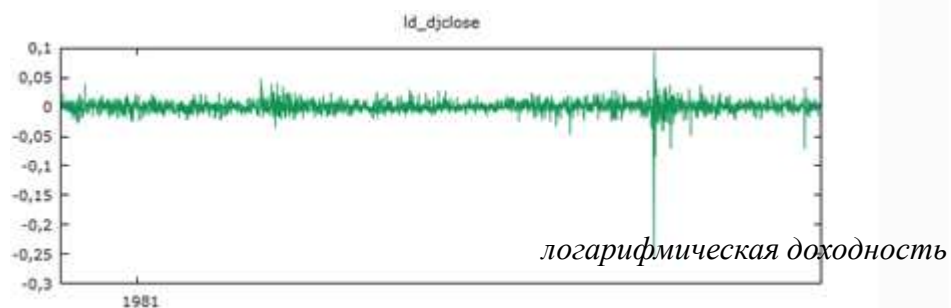
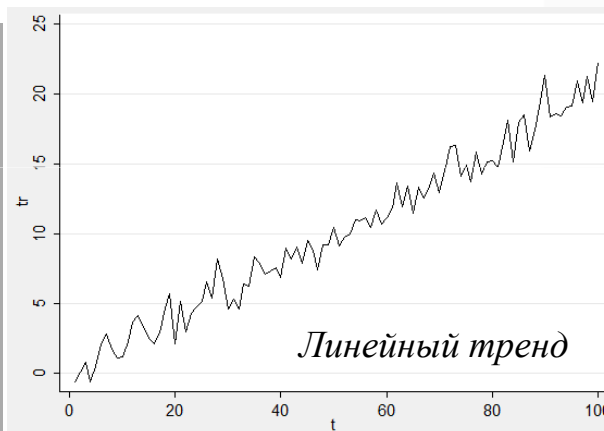
Теоретические

$$\text{Var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

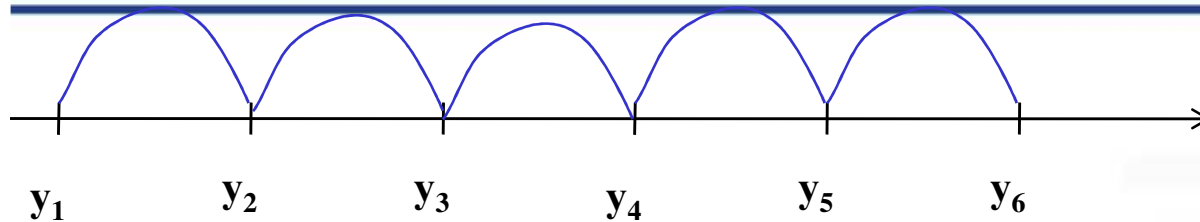
Выборочные

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2$$

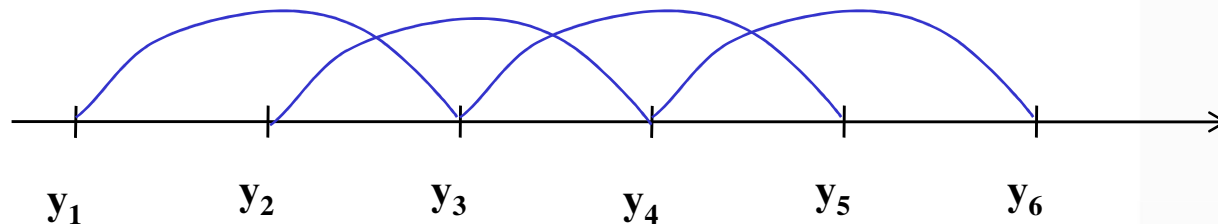
Графическое представление



ВР и их выборочные характеристики: ковариация



γ_1 измеряет линейную взаимосвязь между y_t и y_{t+1}



γ_2 измеряет линейную взаимосвязь между y_t и y_{t+2}

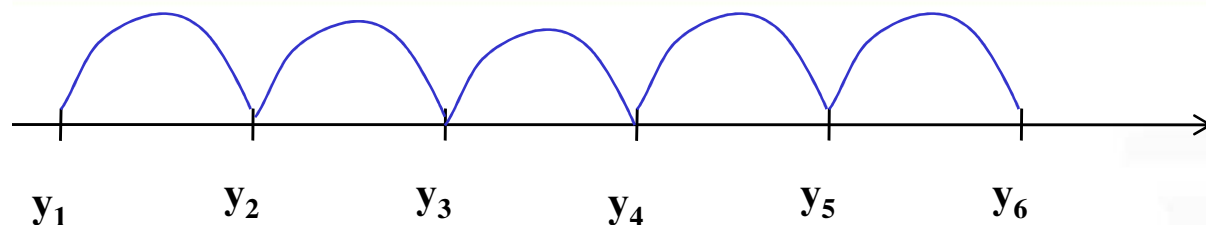
Ковариация наз. **автоковариацией**, т.к. характеризует статистическую взаимосвязь между уровнями одного и того же временного ряда, отстоящих на τ тактов времени.

$$\hat{\gamma}(1) = \frac{1}{T-1} \sum_{t=1}^{T-1} (y_t - \hat{\mu})(y_{t+1} - \hat{\mu})$$

$$\hat{\gamma}(2) = \frac{1}{T-2} \sum_{t=1}^{T-2} (y_t - \hat{\mu})(y_{t+2} - \hat{\mu})$$



ВР и их характеристики: ковариация



γ_1 измеряет линейную взаимосвязь между y_t и $y_{t \pm 1}$

$$\gamma(1) = \gamma_1 = \text{Cov}(Y_t, Y_{t-1}) = E((Y_t - \mu)(Y_{t-1} - \mu))$$

Теоретические

$$\begin{aligned} \gamma(\tau) &= \gamma_\tau = \text{Cov}(Y_t, Y_{t+\tau}) = \\ &= E((Y_t - \mu)(Y_{t+\tau} - \mu)) \end{aligned}$$

$$\hat{\gamma}(0) = \text{Cov}(y_t, y_t) = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2 = \text{Var}(y_t) \quad \text{дисперсия}$$

Выборочные

$$\hat{\gamma}(\tau) = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} (y_t - \hat{\mu})(y_{t+\tau} - \hat{\mu})$$

ВР и их характеристики: автоковариации

$$\gamma(\tau) = \gamma_\tau = \text{Cov}(Y_t, Y_{t-\tau}) = E((Y_t - \mu)(Y_{t-\tau} - \mu))$$

**Выборочная
автоковариация**

$$\hat{\gamma}(\tau) = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} (y_t - \hat{\mu})(y_{t+\tau} - \hat{\mu})$$

$$\hat{\gamma}(\tau) = \hat{\gamma}(-\tau) = \frac{1}{T - \tau} \sum_{t=\tau+1}^T (y_t - \hat{\mu})(y_{t-\tau} - \hat{\mu})$$

Замечание. Возможны альтернативные состоятельные оценки, которые имеют другое скорректированное число степеней свободы.

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} (y_t - \hat{\mu})(y_{t+\tau} - \hat{\mu})$$

Вербик, 2008, с.410

Состоятельная оценка — это точечная оценка, сходящаяся по вероятности к оцениваемому параметру.

Автокорреляционная функция (autocorrelation function, ACF).

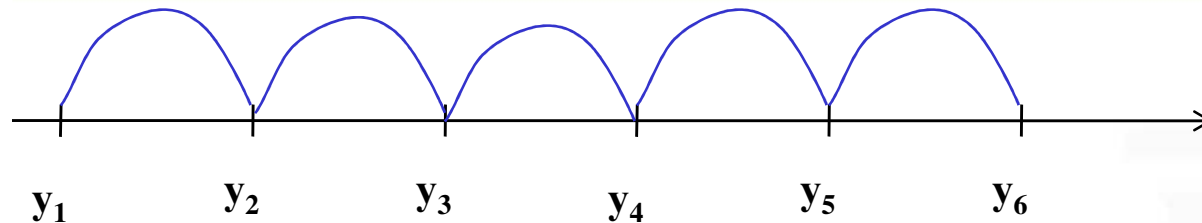
$$\rho(\tau) = \text{Cor}(Y_t, Y_{t-\tau}) = \frac{E((Y_t - \mu)(Y_{t-\tau} - \mu))}{\sqrt{E(Y_t - \mu)^2 E(Y_{t-\tau} - \mu)^2}} = \frac{\gamma(\tau)}{\gamma(0)}$$

$$\hat{\rho}_1 = \text{cor}(y_t, y_{t-1}); \hat{\rho}_2 = \text{cor}(y_t, y_{t-2}); \hat{\rho}_3 = \text{cor}(y_t, y_{t-3});$$

$$\hat{\rho}(\tau) = \frac{\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (y_t - \hat{\mu})(y_{t+\tau} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2}$$

Замечание. Теоретическая ACF и выборочная ACF различаются.
График ACF – **коррелограмма**.

Пример: Автокорреляция 1-го порядка



ρ_1 измеряет линейную взаимосвязь между y_t и $y_{t \pm 1}$

Численность населения России

Рассчитайте коэффициенты корреляции 1, 2, 3-го порядков.

Годы	Все население, млн. человек
2015	146,3
2016	146,5
2017	146,8
2018	146,9

$$\hat{\rho}_1 = cor(y_t, y_{t-1});$$

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

$$\hat{\mu} = \bar{y}_t = 146.6; \sigma^2 = 0.09$$

$$\hat{\rho}_1 = \frac{\frac{1}{3} \sum_{t=2}^4 (y_t - 146.6)(y_{t-1} - 146.6)}{0.09} =$$

$$\rho_1 = 0,7$$

t	Yt	Yt-1	Yt+1
2015	146,3		146,5
2016	146,5	146,3	146,8
2017	146,8	146,5	146,9
2018	146,9	146,8	

Аналогично

$\rho_2, \rho_3, \rho_4, \dots$

$$\hat{\rho}_2 = cor(y_t, y_{t-2}); \hat{\rho}_3 = cor(y_t, y_{t-3});$$

$$\hat{\rho}(\tau) = \frac{\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (y_t - \hat{\mu})(y_{t+\tau} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2}$$

Автокорреляция

Автокорреляционная функция (autocorrelation function, ACF) – зависимость значений автокорреляций (серийных автокорреляций) от лага τ

ACF	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
	0,72	0,39	-0,02	-0,33	-0,46

График

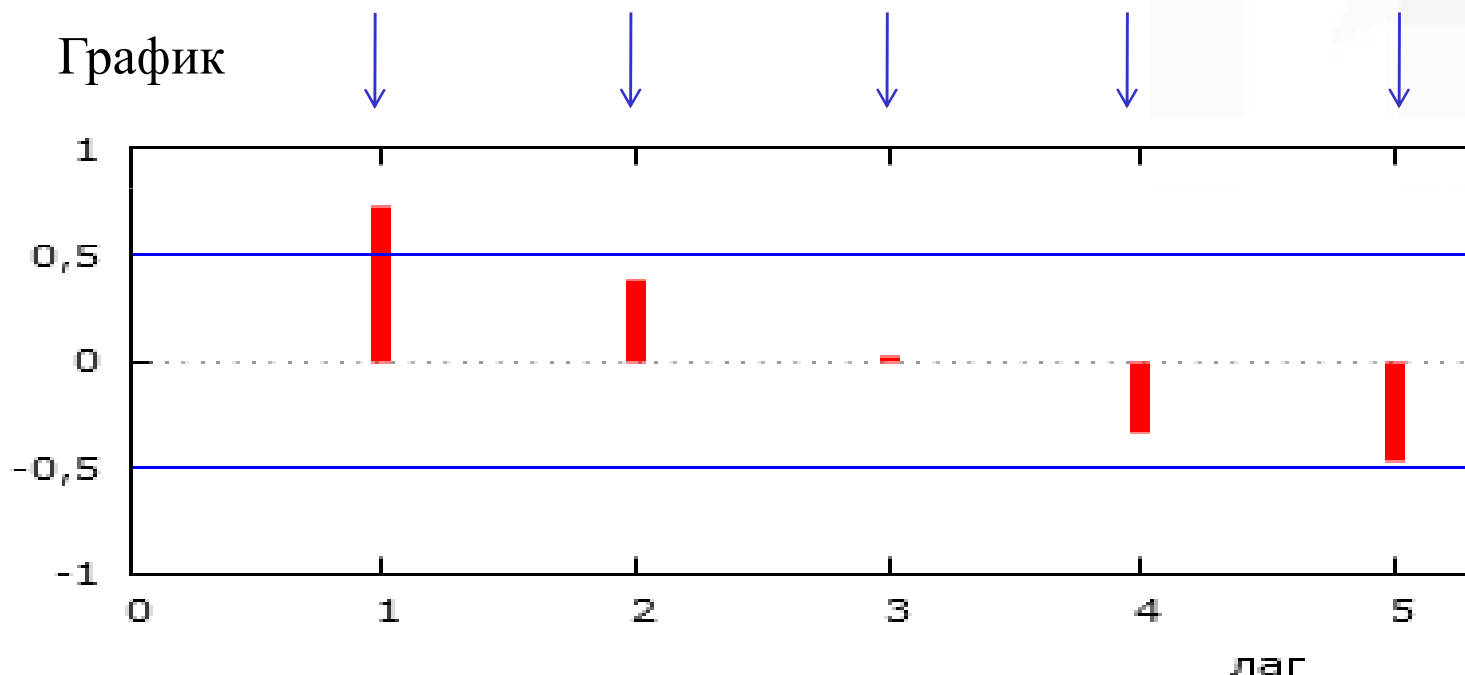
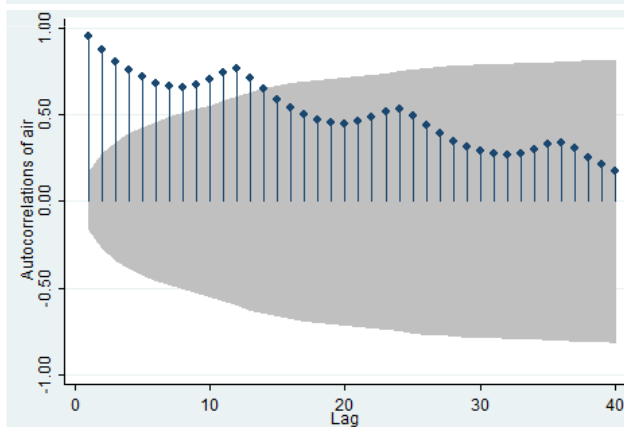
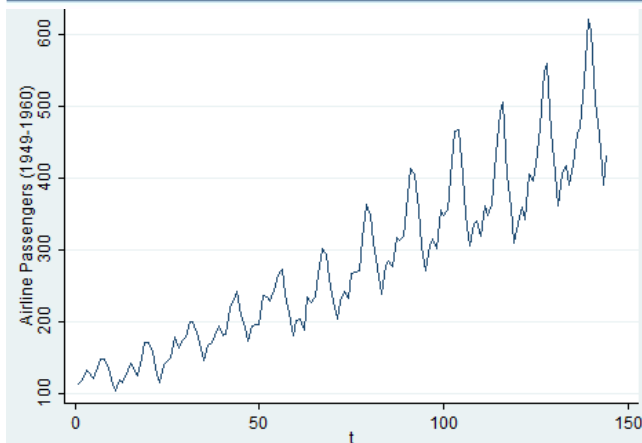


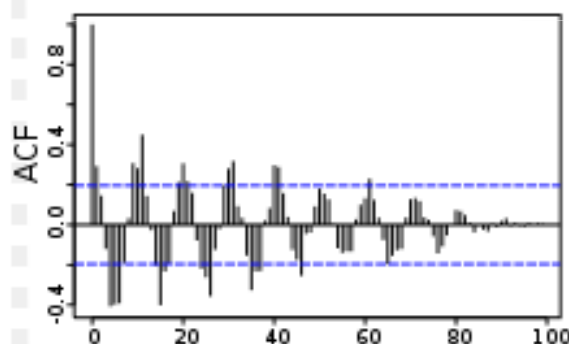
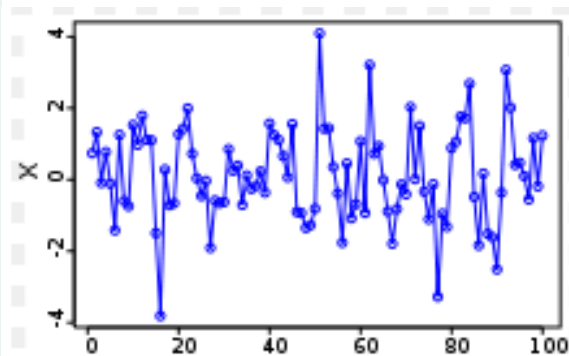
График ACF – коррелограмма



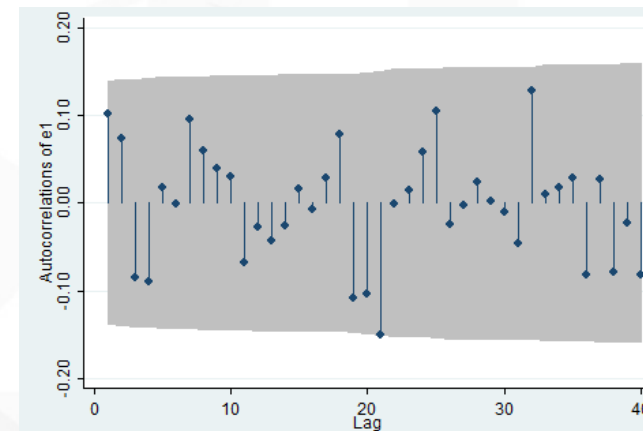
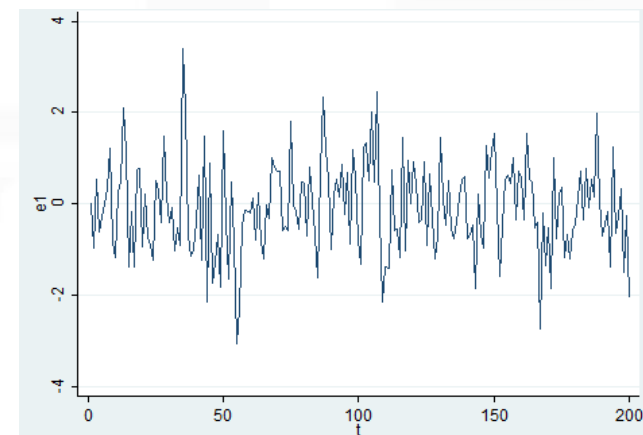
Автокорреляционная функция: примеры



Выявление сезонности



Данные со скрытой синусоидой.
Выявление периодичности.

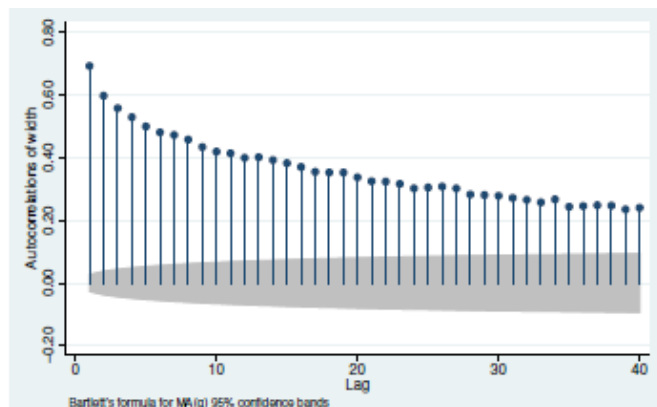
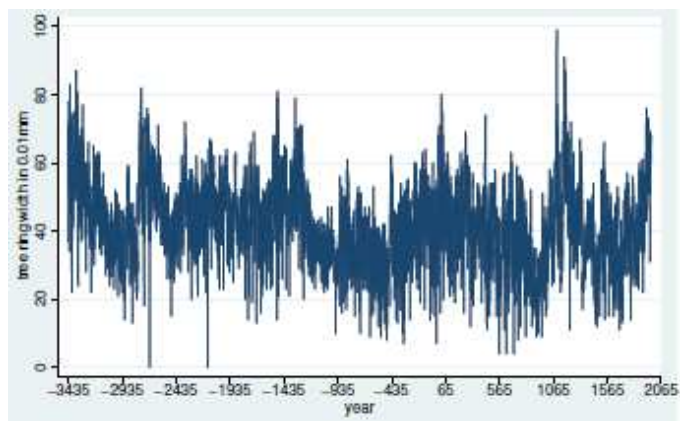


белый шум



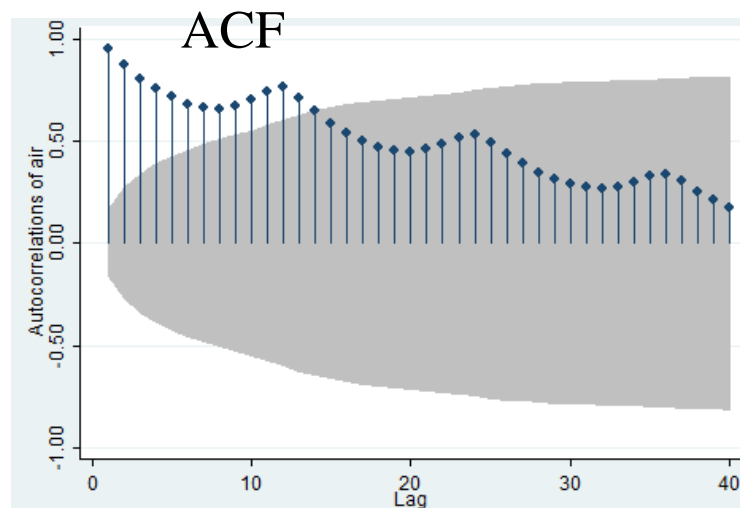
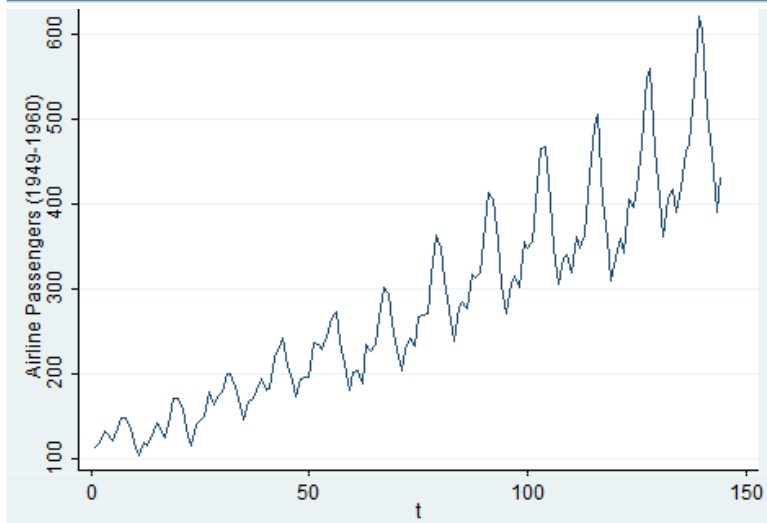
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ «ВШЭ»
VNIIS

Автокорреляционная функция: примеры



Процесс с длинной памятью

Автокорреляционная функция



Свойства:

1. ACF безразмерна
2. $|\rho(\tau)| \leq 1$
3. $\rho(\tau) = \rho(-\tau)$
4. Если y_t – белый шум, то при $n \rightarrow \infty$ выборочная ACF имеет нормальное распределение:

$$\hat{\rho}_\tau \approx N(0, \frac{1}{N})$$

(используется для проверки значимости ACF)

Простейшие примеры временных рядов: белый шум

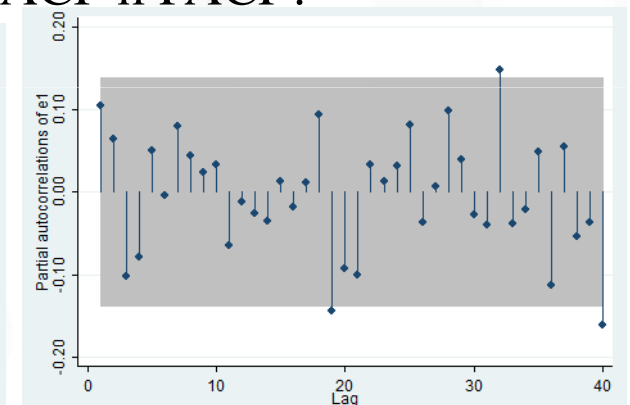
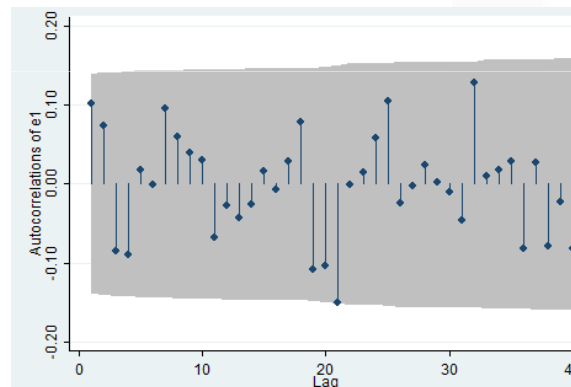
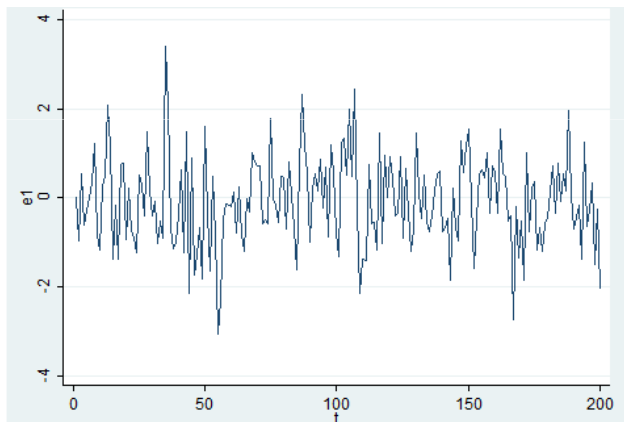
Белый шум (white noise) - процесс ε_t , удовлетворяющий условиям теоремы Гаусса-Маркова:

$$E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2$$

$$\text{Cov}(\varepsilon_p, \varepsilon_{t-s}) = 0 \ (s \neq 0)$$

Если ε_t распределены в совокупности нормально, то процесс наз. **гауссовым белым шумом**.

Какой вид имеют ACF и PACF?



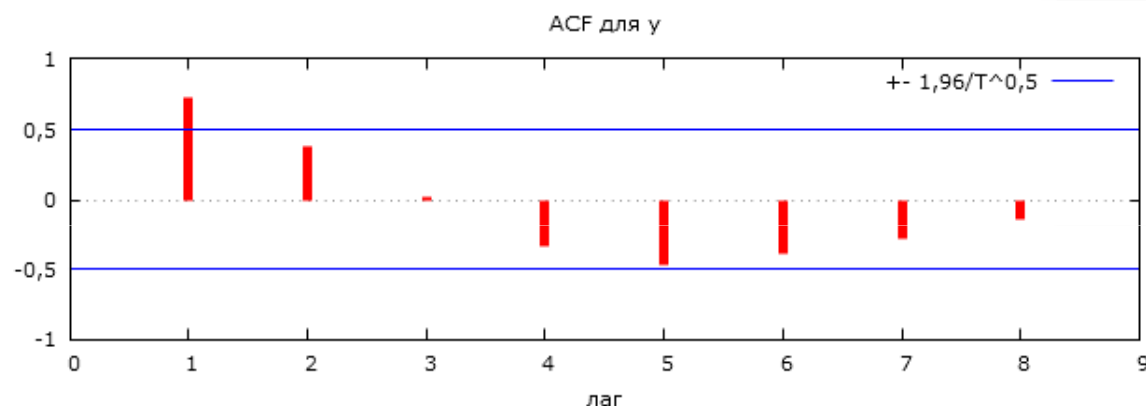
Свойство: Если y_t – белый шум, то при $n \rightarrow \infty$ выборочная ACF имеет нормальное распределение: $\hat{\rho}_\tau \approx N(0, \frac{1}{N})$

Интерпретация: 95% пиков в ACF находятся в пределах $\pm 1,96/\sqrt{T}$

Автокорреляционная функция

Свойство 4 используется для проверки значимости АСФ.

Пример. Пусть $T=9$. Если ряд является БШ, то 95% значений АСФ должны лежать в интервале $\pm 1,96/\sqrt{9} \sim \pm 0,65$

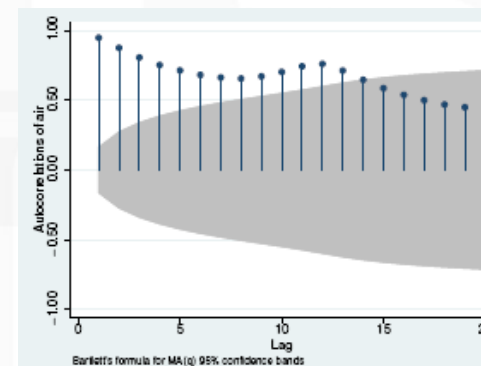


Вывод: данный ряд не является реализацией процесса БШ

В Stata для проверки значимости АСФ используется аппроксимация Бартлетта

Bartlett's formula

$$\text{Var}(\hat{\rho}_v) = \begin{cases} 1/n & v = 1 \\ \frac{1}{n} \left\{ 1 + 2 \sum_{i=1}^{v-1} \hat{\rho}^2(i) \right\} & v > 1 \end{cases}$$



$$\hat{\rho}(\tau) = \frac{\frac{1}{T-\tau} \sum_{t=\tau+1}^T (y_t - \hat{\mu})(y_{t-\tau} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2}$$

Пример

t	Yt	Yt-1	Yt-2	Yt-3
2015	146,3			
2016	146,5	146,3		
2017	146,8	146,5	146,3	
2018	146,9	146,8	146,5	146,3

Пример. Если T=4, то для вычисления $\rho(3)$ исп-ся 1 наблюдение

- по ряду длиной T можно вычислить $\rho(\tau)$ до порядка (T-1).
- «дальние» $\rho(\tau)$ вычисляются неточно: с ростом τ количество наблюдений, по которым считаются автокорреляции уменьшается (для расчета $\rho(T-1)$ используется 1 наблюдение).
- при анализе ACF следует принимать во внимание «ближние» $\rho(\tau)$ (первые $[T/5]$ автокорреляций)

Суслов

Частная автокорреляционная функция (partial autocorrelation function, PACF) – аналог частной корреляции. (Айвазян)

-значение $\rho_{part}(k)$ на лаге k определяется как значение коэффициента корреляции между случайными величинами Y_t и Y_{t+k} , очищенными от влияния случайных величин $Y_{t+1}, \dots, Y_{t+k-1}$.

$$\rho_{част}(1) = \text{Cor}(Y_t, Y_{t+1}) = \rho(1)$$

$$\rho_{част}(2) = \text{Cor}(Y_t, Y_{t+2} \mid Y_{t+1} = \mu) = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}$$

$$\rho_{част}(3) = \dots \quad (\text{выписать самостоятельно})$$

Опр. 1.3. Временной ряд наз. *стационарным (слабо стационарным) (weak stationary)* (в широком смысле), если он обладает постоянной средней и дисперсией, а ковариация зависит только от временного интервала между отдельными наблюдениями.

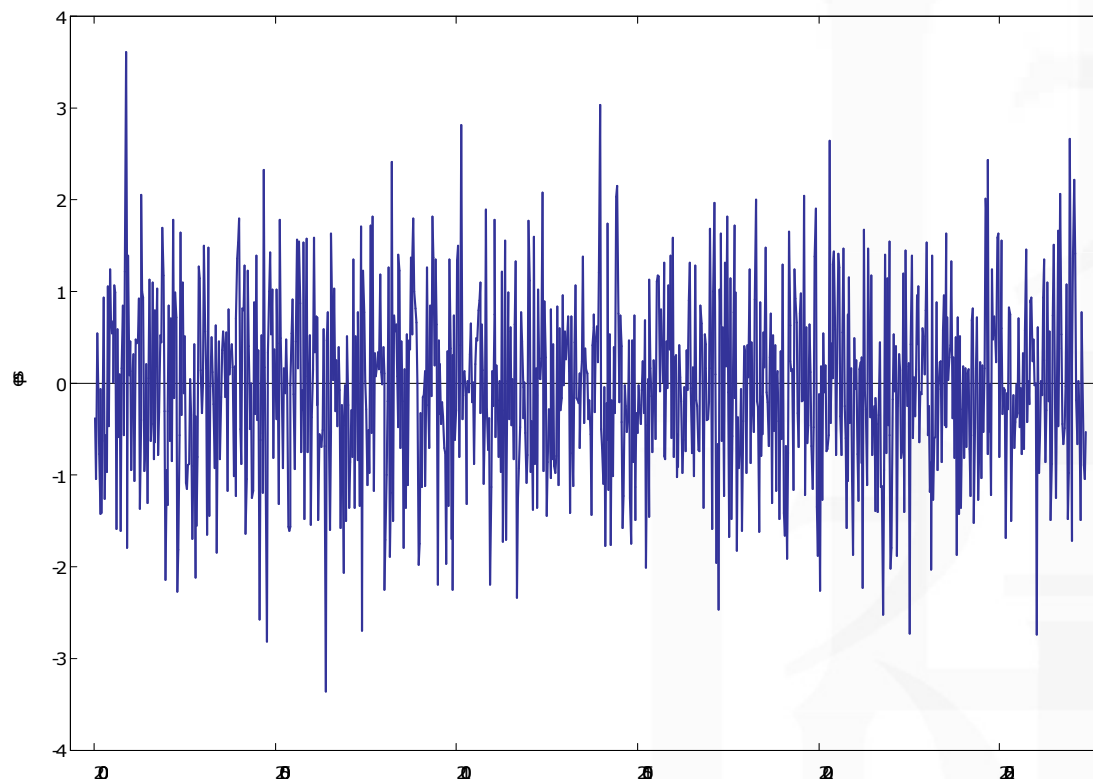
$$1. E(Y_t) = \mu$$

$$2. V(Y_t) = \sigma^2$$

$$3. \text{Cov}(Y_t, Y_{t+\tau}) = E[(Y_t - \mu)(Y_{t+\tau} - \mu)] = \gamma(\tau)$$

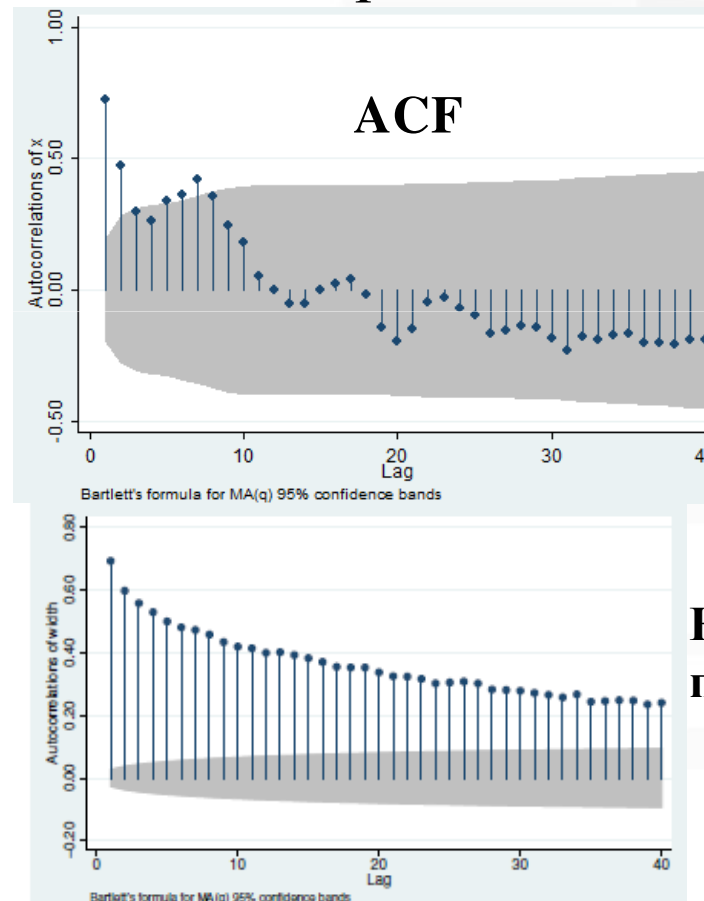
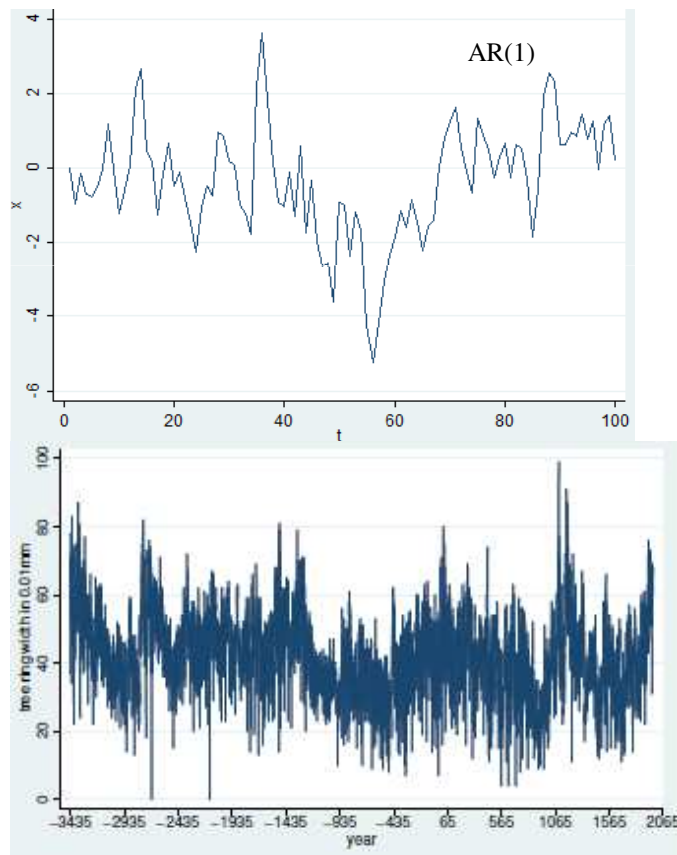
Стационарность ВР

Общая идея: стационарность \rightarrow поведения ряда в настоящем и будущем совпадает с его поведением в прошлом.



Стационарные ВР: АСФ и РАСФ

Для стационарных ВР коррелограмма (график АСФ и РАСФ)
«быстро убывает» после нескольких первых значений.



**Стационарный
процесс**

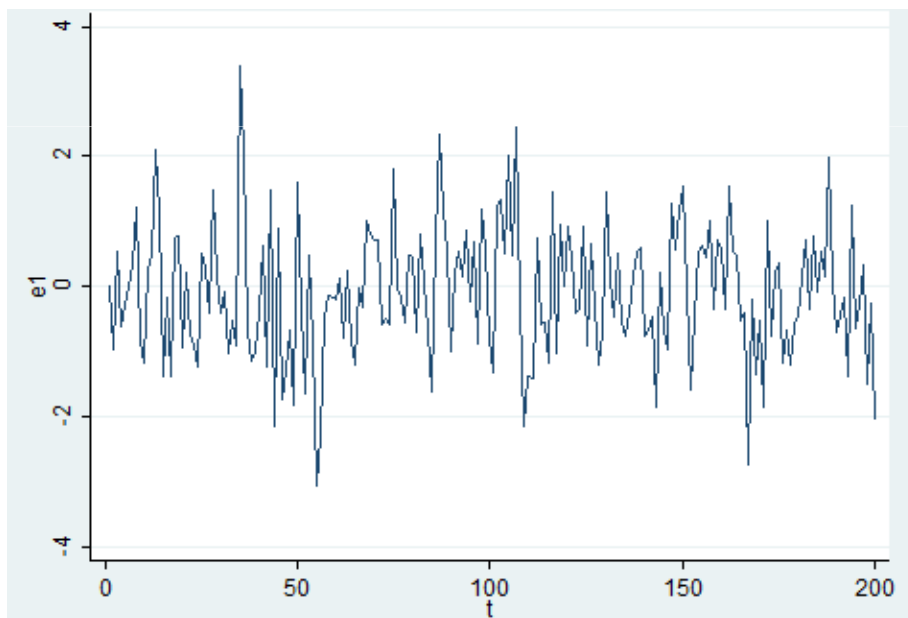
**Нестационарный
процесс**

Простейшие примеры временных рядов

1. **Белый шум (white noise)** - процесс ε_t , удовлетворяющий условиям теоремы Гаусса-Маркова:

$$E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2$$

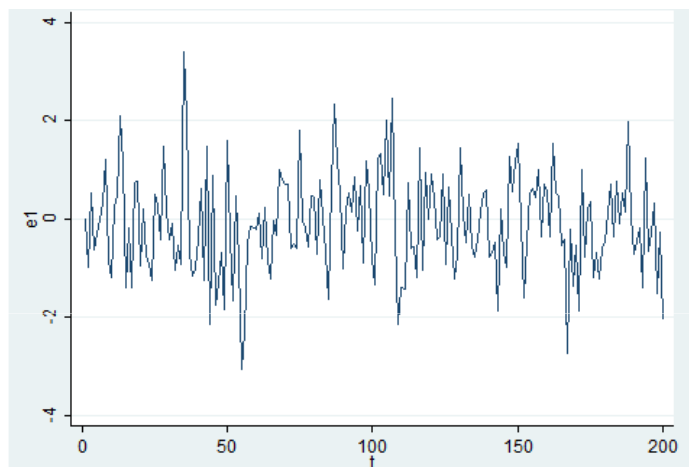
$$\text{Cov}(\varepsilon_p, \varepsilon_{t-s}) = 0 \ (s \neq 0)$$



Если ε_t распределены в совокупности нормально, то процесс наз. **гауссовым белым шумом**.

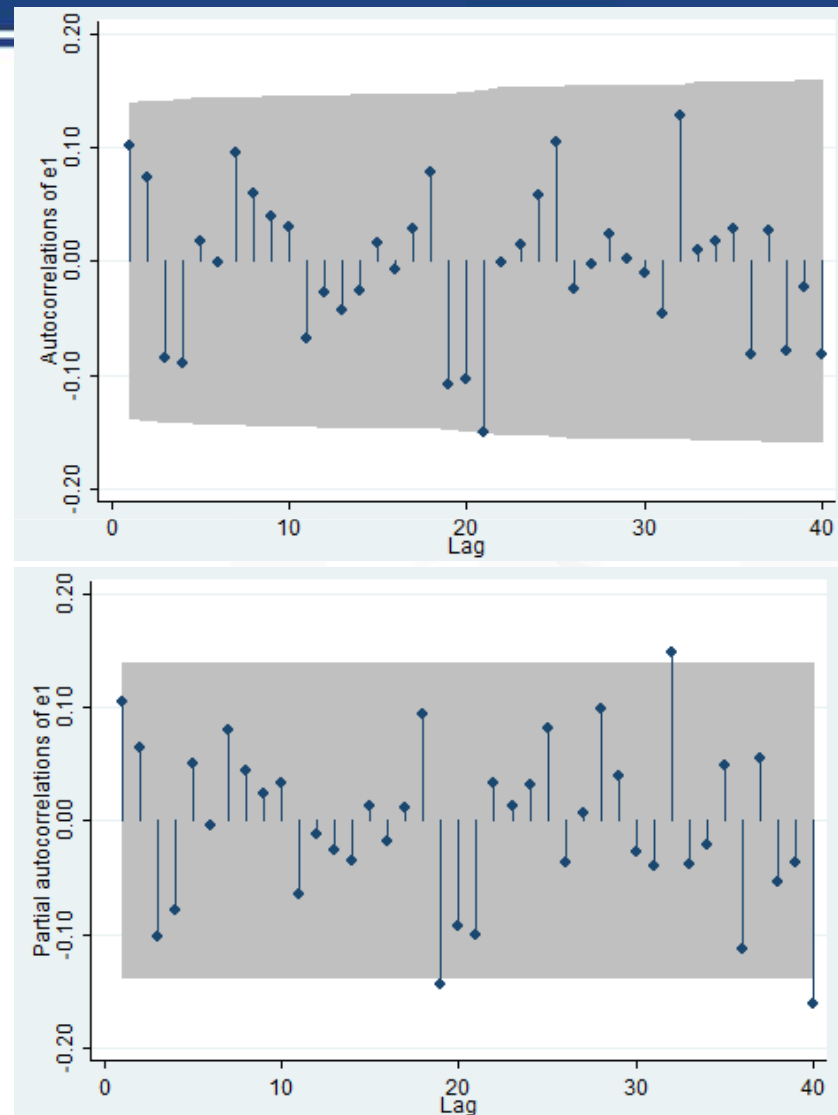
Какой вид имеют ACF и PACF?

Белый шум



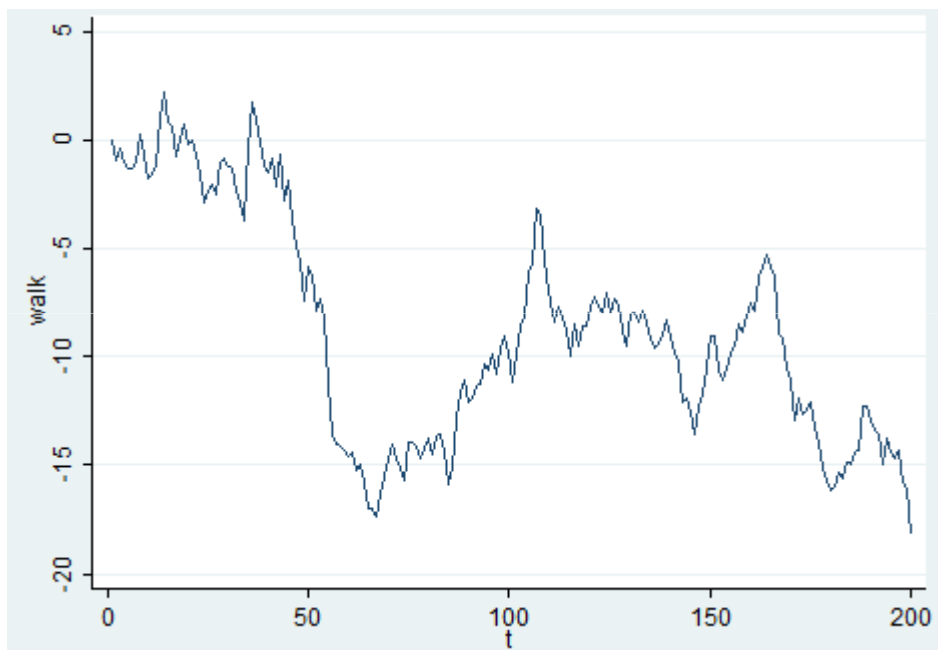
$$E(\varepsilon_t) = 0,$$

$$V(\varepsilon_t) = \sigma^2$$



Простейшие примеры временных рядов

2. случайное блуждание (random walk) $Y_t = Y_{t-1} + \varepsilon_t$ (броуновское движение)



Стационарен или нет?

$$E(Y_t) =$$

$$V(Y_t) =$$



Простейшие примеры временных рядов

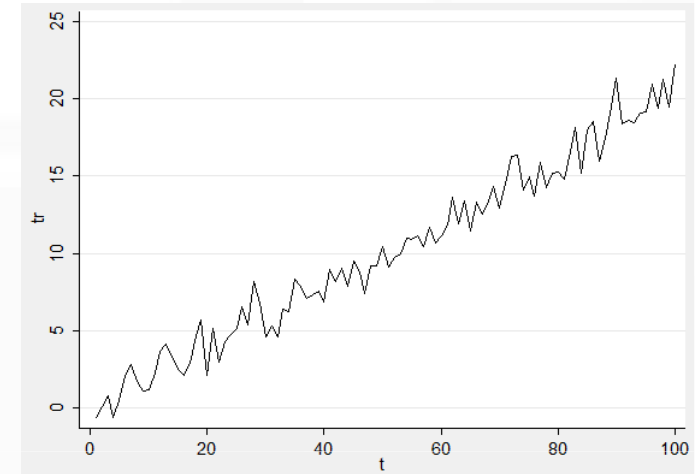
3. Процесс с линейным трендом

$$Z_t = \alpha + \beta t + \varepsilon_t$$

Стационарен или нет?

$$E(Z_t) =$$

$$V(Z_t) =$$



4. случайное блуждание с дрейфом (random walk with drift)

$$X_t = \mu + X_{t-1} + \varepsilon_t$$

(броуновское движение)

Стационарен или нет?

$$E(X_t) =$$

$$V(X_t) =$$

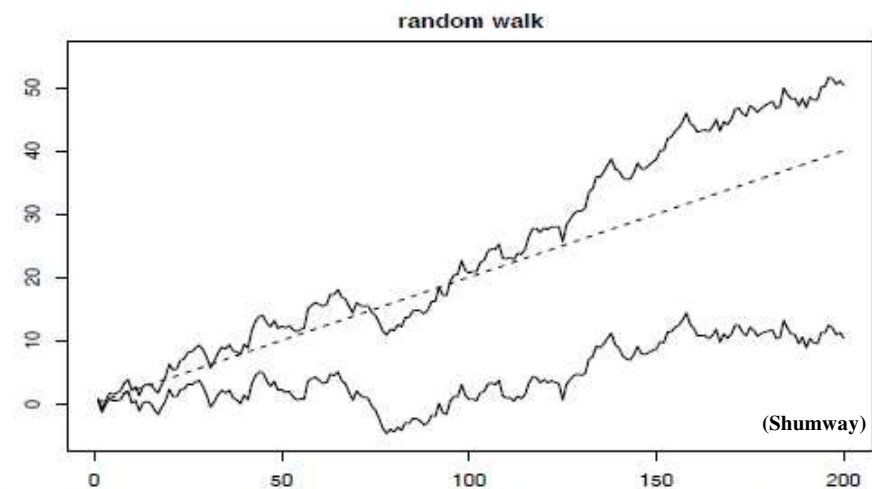


Fig. 1.10. Random walk, $\sigma_w = 1$, with drift $\delta = .2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and a straight line with slope .2 (dashed line).

5. Случайный процесс как скользящее среднее белого шума

$$Z_t = 1/3(\varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1})$$

Стационарен или нет?

$$E(Z_t) =$$

$$V(Z_t) =$$

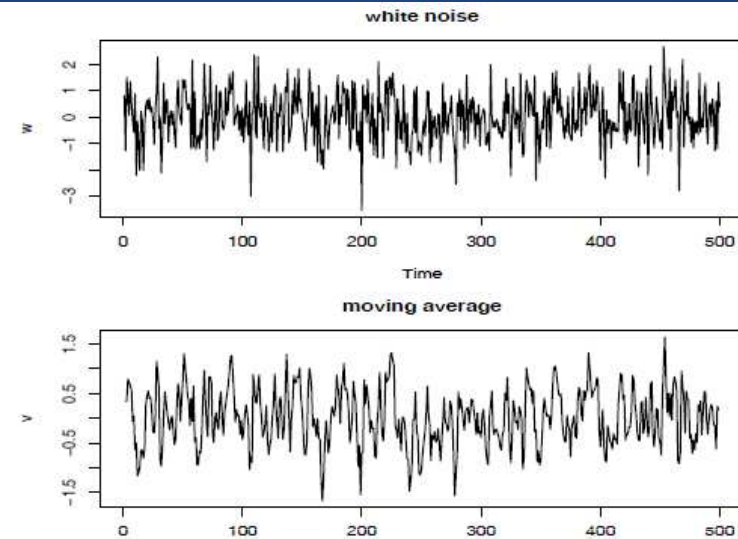


Fig. 1.8. Gaussian white noise series (top) and three-point moving average of the Gaussian white noise series (bottom).

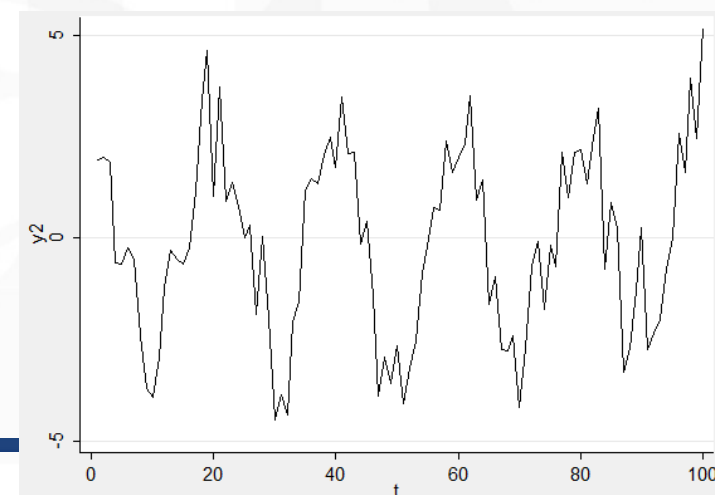
6. Случайный процесс с периодической составляющей

$$X_t = 3\cos(2\pi t/20) + \varepsilon_t$$

Стационарен или нет?

$$E(X_t) =$$

$$V(X_t) =$$





Лаговый и разностный оператор

Лаговый оператор (Lag operator) :
(оператор запаздывания)

Обеспечивает сжатую запись уравнений.

Свойства:

$$1) LC = C,$$

$$2)(L^j + L^i)y_t = y_{t-j} + y_{t-i},$$

$$3)L^j L^i y_t = y_{t-j-i},$$

$$4)L^{-i} y_t = y_{t+i},$$

$$5)|a| < 1 \rightarrow (1 + aL + a^2 L^2 + a^3 L^3 + \dots)y_t = \frac{y_t}{1-aL}$$

$$Ly_t = y_{t-1}, \quad t = 2, \dots, n$$

$$L^2 y_t = y_{t-2}, L^4 y_t = y_{t-4},$$

$$L^k y_t = y_{t-k},$$

Stata

LY_t=L.Y

L² Y_t=L2.Y



Лаговый и разностный оператор

Разностный оператор (Difference operator):

$$\Delta y_t = y_t - y_{t-1} = y_t - L y_t = (1 - L) y_t$$

$$\Delta^2 y_t = \Delta \Delta y_t = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

$$\Delta^3 y_t =$$

! Записать самостоятельно

Stata

$$\Delta Y_t = D.Y$$

$$\Delta^2 Y_t = D2.Y$$

Пример y_t 1 2 3 4 5

Найти $L y_t$, $L^2 y_t$, Δy_t , $\Delta^2 y_t$



Разностный оператор: частный случай

Разностный оператор (Difference operator): Fractional differencing

Если d - не целое число?

$$\Delta^d y_t = (1 - L)^d y_t = \dots$$

$$(1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k =$$
$$= 1 - dL + \frac{d(d-1)}{2!} L^2 - \frac{d(d-1)(d-2)}{3!} L^3 + \dots$$

Пример. $d=0.5 \rightarrow (1 - L)^{0.5} =$

$$(1 - L)^{0.5} y_t =$$

Задание. $d=0.1 \rightarrow (1 - L)^{0.1} y_t =$ записать самостоятельно

Ряды Маклорена некоторых функций
Биномиальное разложение:

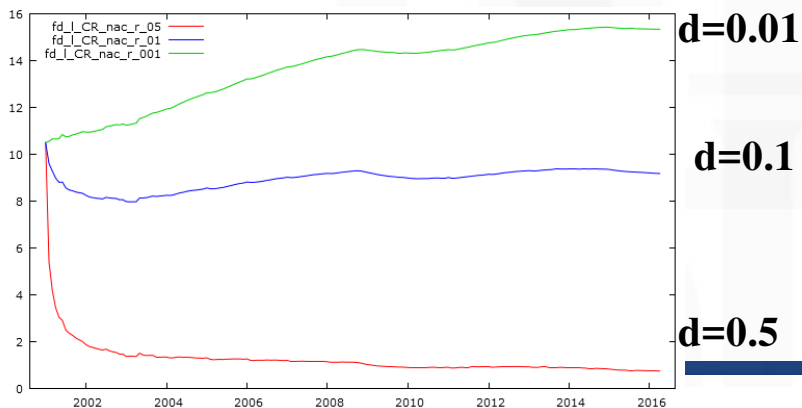
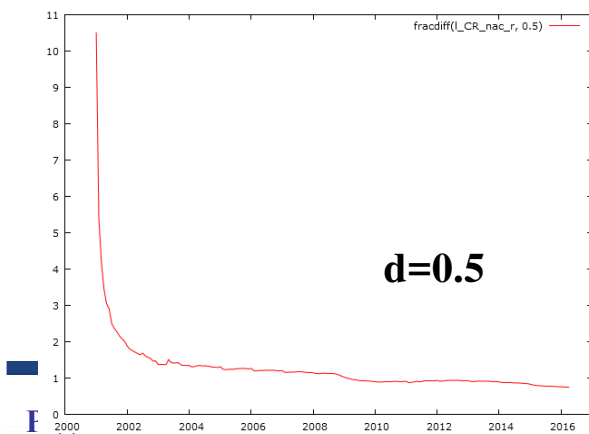
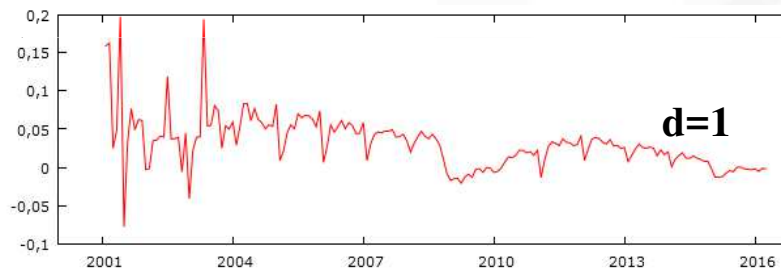
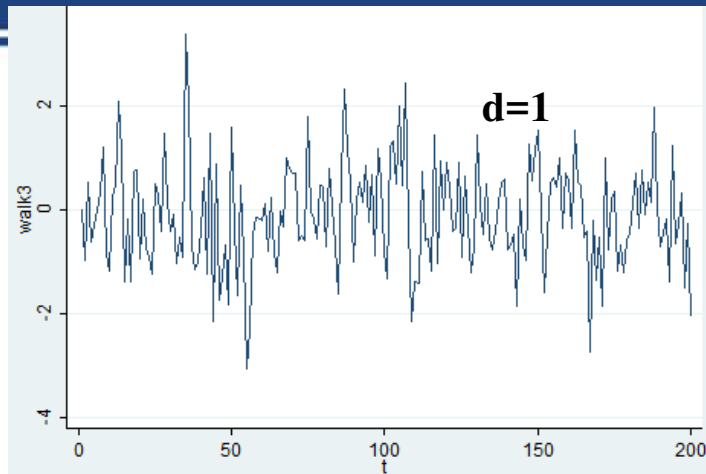
$$(1 + x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k$$
$$= 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \dots$$
$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)(\alpha-2) \cdots (\alpha-k+1)}{k!}$$

https://en.wikipedia.org/wiki/Binomial_series

**!Дробное дифференцирование
используется в ARFIMA
моделях**



Разностный оператор: преобразования рядов





Разностный оператор: реализация в пакетах

Stata

Time Series Operators

- L. Lag y_{t-1} (L1. means the same thing)
- L2. 2-period lag y_{t-2} (similarly, L3., etc. L(1/4). means L1. through L4.)
- F. Lead y_{t+1} (F1. means the same thing)
- F2. 2-period lead y_{t+2} (similarly, F3., etc.)
- D. Difference $y_t - y_{t-1}$ (D1. means the same thing)
- D2. Second difference $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ (similarly, D3., etc.)

list year cpi L.cpi L2.cpi L3.cpi

	year	cpi	L. cpi	L2. cpi	L3. cpi
1.	1989	124	.	.	.
2.	1990	130.7	124	.	.
3.	1991	136.2	130.7	124	.
4.	1992	140.3	136.2	130.7	124
5.	1993	144.5	140.3	136.2	130.7
6.	1994	148.2	144.5	140.3	136.2
7.	1995	152.4	148.2	144.5	140.3
8.	1996	156.9	152.4	148.2	144.5

list year cpi D.cpi D2.cpi D3.cpi

	year	cpi	D. cpi	D2. cpi	D3. cpi
1.	1989	124	.	.	.
2.	1990	130.7	6.699997	.	.
3.	1991	136.2	5.5	-1.199997	.
4.	1992	140.3	4.100006	-1.399994	-.1999969
5.	1993	144.5	4.199997	.0999908	1.499985
6.	1994	148.2	3.699997	-.5	-.5999908
7.	1995	152.4	4.199997	.5	1
8.	1996	156.9	4.5	.3000031	-.1999969

list year cpi F.cpi F2.cpi F3.cpi

	year	cpi	F. cpi	F2. cpi	F3. cpi
1.	1989	124	130.7	136.2	140.3
2.	1990	130.7	136.2	140.3	144.5
3.	1991	136.2	140.3	144.5	148.2
4.	1992	140.3	144.5	148.2	152.4
5.	1993	144.5	148.2	152.4	156.9
6.	1994	148.2	152.4	156.9	.
7.	1995	152.4	156.9	.	.
8.	1996	156.9	.	.	.

Лаговый многочлен (многочлен от оператора сдвига) – линейная комбинация лагов в записи моделей ВР.

$$C(L) = \sum_{j=0}^{\infty} c_j L^j$$

$C(L)$ преобразует X_t в процесс Y_t такой что:

$$Y_t = C(L)X_t = \sum_{j=0}^{\infty} c_j L^j X_t = \sum_{j=0}^{\infty} c_j X_{t-j}$$