АВТОРЕГРЕССИОННЫЕ МОДЕЛИ

1. Авторегрессионные (AR-) модели
2. Стационарность временных рядов. Тест Дики-Фуллера
3. Авторегрессионные модели с условной гетероскедастичнонстью
4. Взаимосвязанные временные ряды

Автокорреляцией уровней ряда называется корреляционная зависимость между последовательными уровнями временного ряда. Для измерения автокорреляции уровней ряда используется коэффициент автокорреляции.

Автокорреляция обычно возникает при изучении временных рядов (когда номер наблюдения соответствует моменту времени).

Коэффициент автокорреляции r_t порядка t определяется как коэффициент корреляции между рядами y_t и y_{t-i} Число периодов i, по которым рассчитывается коэффициент автокорреляции, называется ***лагом***.

Коэффициенты автокорреляции разных порядков обозначают как r_k, где $k = 1$, 2, ... указывает на порядок коэффициента автокорреляции.

Автокорреляционной функцией (ACF) временного ряда называется последовательность коэффициентов автокорреляции первого, второго и более высоких порядков. ***Коррелограммой*** называется график зависимости ее значений от величины лага (порядка коэффициента автокорреляции).

Автокорреляционную функцию обычно используют для выявления во временном ряде наличия или отсутствия трендовой и сезонной компонент:

- если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд *содержит только тенденцию*;
- если наиболее высоким оказался коэффициент автокорреляции порядка $t > 1$, то ряд содержит сезонные колебания с периодом t;
- если ни один из коэффициентов автокорреляции не является значимым, то либо ряд не содержит тенденции и сезонных колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Авторегрессионная (*AR*-) модель (англ. autoregressive model) – модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда.

Авторегрессионный процесс порядка p ($AR(p)$-процесс):

$$Y_t = c + \sum_{i=1}^{p} a_i Y_{t-i} + \varepsilon_t$$

где $a_1, a_2, \dots, a_\mathrm{p}$ – параметры модели (коэффициенты авторегрессии);

c – константа (иногда прнимается равной нулю);

ε_t – белый шум.

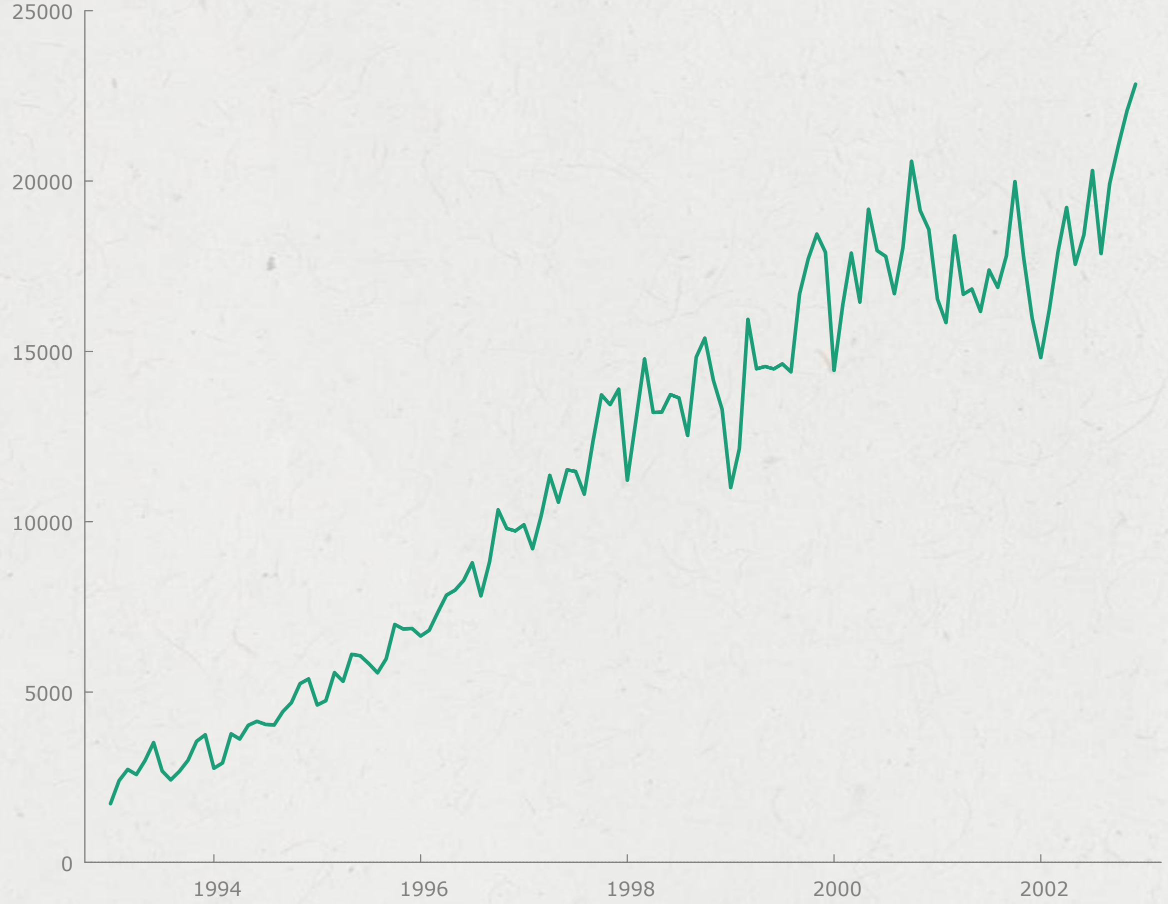
Авторегрессионный процесс первого порядка ($AR(1)$-процесс):

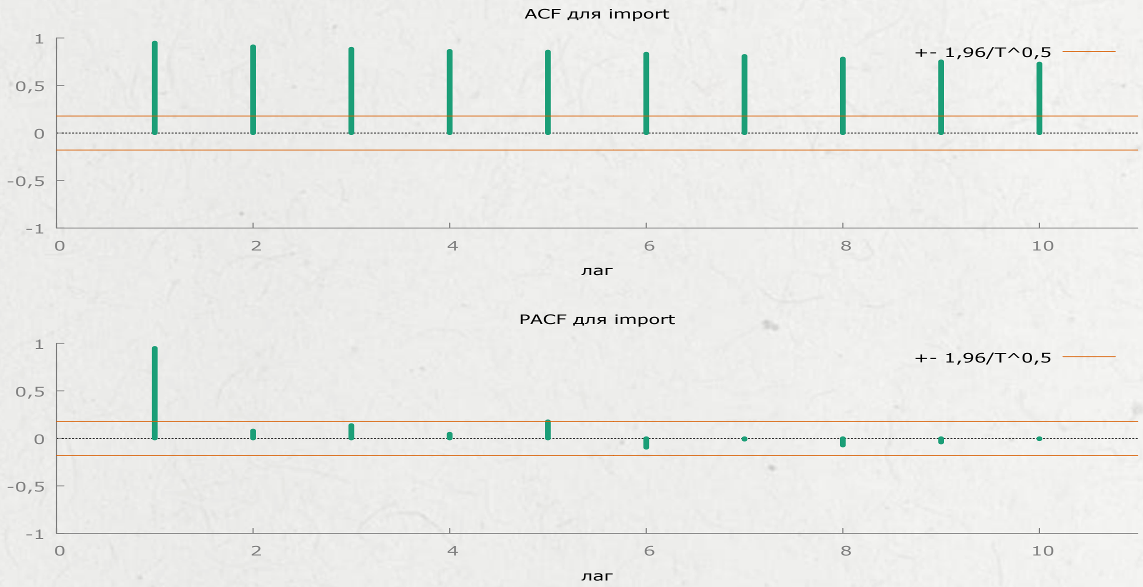
$$Y_t = c + rY_{t-1} + \varepsilon_t$$

Авторегрессионный процесс второго порядка – процесс Юла – ($AR(2)$-процесс):

$$Y_t = c + a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$$

Пример авторегрессионного процесса первого порядка ($AR(1)$-процесс):





Автокорреляционная функция для import
***, **, * обозначает значимость на 1%, 5%, 10% уровне
использованы стандартные ошибки 1/T^0,5

Лаг	ACF		PACF		Q-стат.	[p-значение]
1	0,9503	***	0,9503	***	111,1098	[0,000]
2	0,9110	***	0,0807		214,0668	[0,000]
3	0,8858	***	0,1379		312,2509	[0,000]
4	0,8634	***	0,0477		406,3233	[0,000]
5	0,8557	***	0,1780	*	499,5305	[0,000]
6	0,8342	***	-0,0971		588,8986	[0,000]
7	0,8098	***	-0,0122		673,8670	[0,000]
8	0,7825	***	-0,0738		753,9089	[0,000]
9	0,7528	***	-0,0422		828,6585	[0,000]
10	0,7298	***	-0,0077		899,5487	[0,000]

Часто экономические показатели, представленные временными рядами, имеют настолько сложную структуру, что моделирование таких рядов путем построения моделей тренда, сезонности и применение других традиционных подходов не приводят к удовлетворительным результатам. Ряд остатков часто имеет статистические закономерности, которые можно моделировать.

Будем рассматривать класс стационарных временных рядов. Задача состоит в *построении модели для случайных остатков временного ряда и прогнозирования его значений*.

Стационарными называют случайные процессы, статистические характеристики которых (среднее, дисперсия и ковариация) не изменяются во времени. Если же хотя бы одно из этих свойств непостоянно, то процесс будет нестационарным. Очевидно, что при наличии тренда нельзя утверждать, что среднее значение, дисперсия и автоковариация ряда не зависят от времени, а значит, ряд нестационарен (для стационарного ряда недопустимы тренд и циклическая компонента).

Для обнаружения нестационарности временных рядов используется тест Дики - Фуллера (*DF*-тест).

Рассмотрим модель *AR*(1): $y_t = \alpha y_{t-1} + \varepsilon_t$

где ε_t - белый шум.

Вычтем y_{t-1} из обеих частей уравнения для y_t, (взятие первой разности):

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t$$

где $\Delta y_t = y_t - y_{t-1}$,

$\beta = \alpha - 1$

Временной ряд имеет единичный корень, или порядок интеграции один, если его первые разности образуют стационарный ряд.

Нулевая гипотеза H_0: β = 0 (временной ряд нестационарен) равнозначна H_0: $\alpha = 1$. Это означает, что процесс интегрирован в первой степени: $y_t \sim I(1)$;

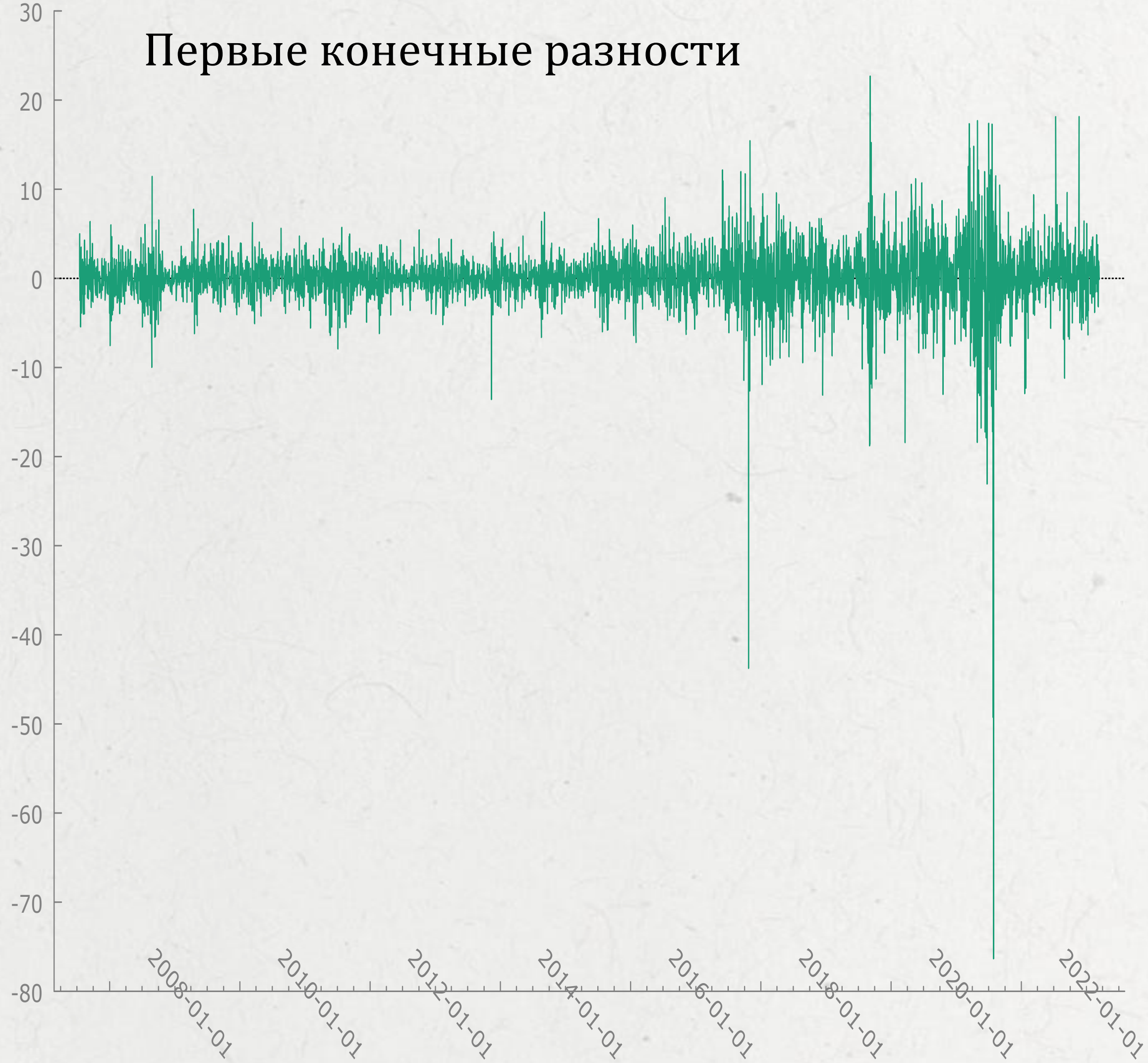
Альтернативная гипотеза H_1: β < 0 (временной ряд стационарен), равнозначна H_1: $\alpha < 1$. В этом случае $y_t \sim I(0)$. Отрицательная значимость параметра β означает отсутствие интеграции процесса.

Авторегрессионные модели с условной гетероскедастичнонстью (англ. ***ARCH*** – *AutoRegressive Conditional Heteroscedasticity*) – применяемая в эконометрике модель для анализа временных рядов, у которых условная (по прошлым значениям ряда) дисперсия ряда зависит от прошлых значений ряда, прошлых значений этих дисперсий и иных факторов.

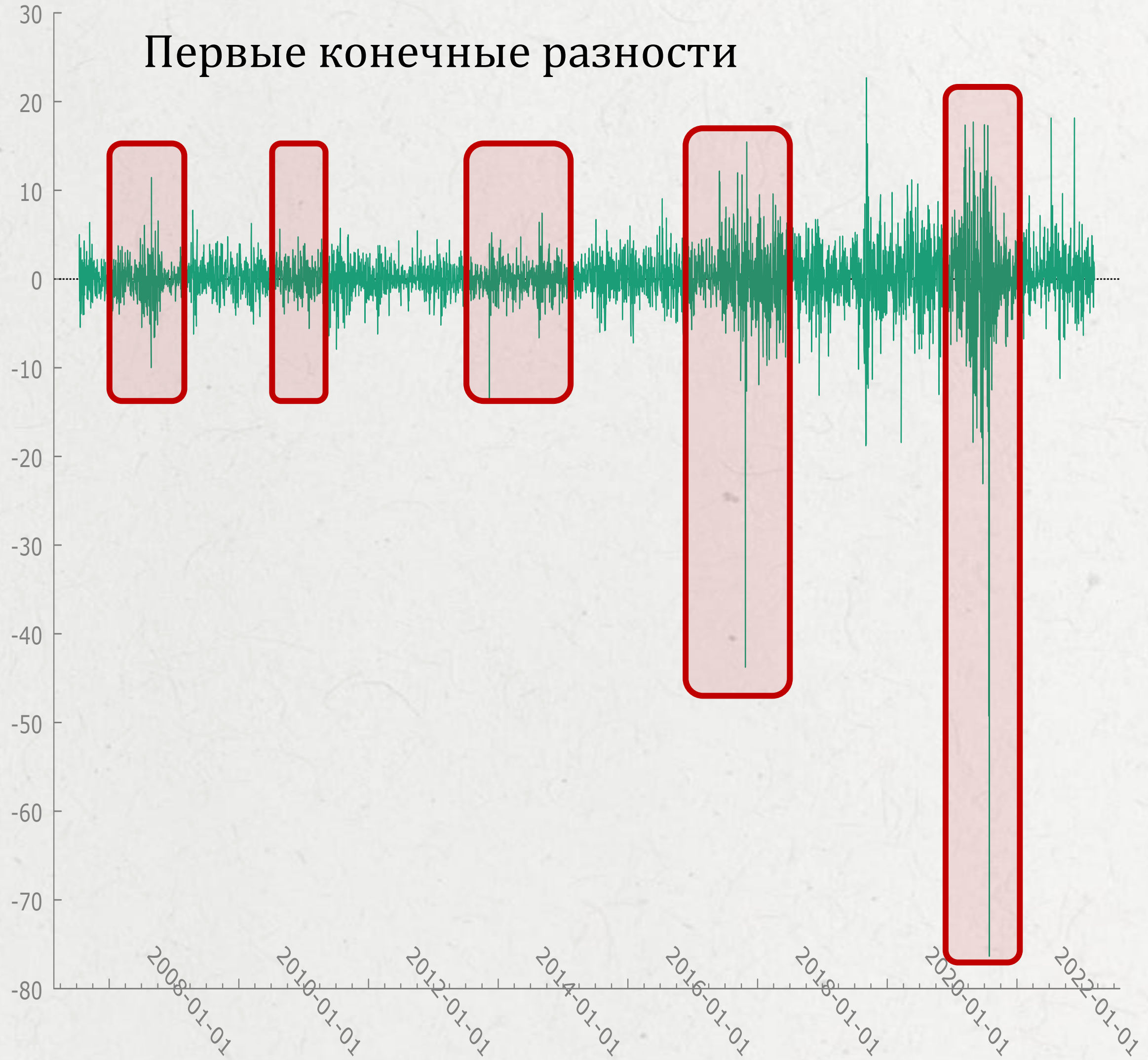
Данные модели предназначены для «объяснения» кластеризации волатильности на финансовых рынках, когда периоды высокой волатильности длятся некоторое время, сменяясь затем периодами низкой волатильности, причём среднюю (долгосрочную, безусловную) волатильность можно считать относительно стабильной.

Волатильность (volatility – изменчивость) – статистический показатель, характеризующий степень вариабельности цены актива во времени

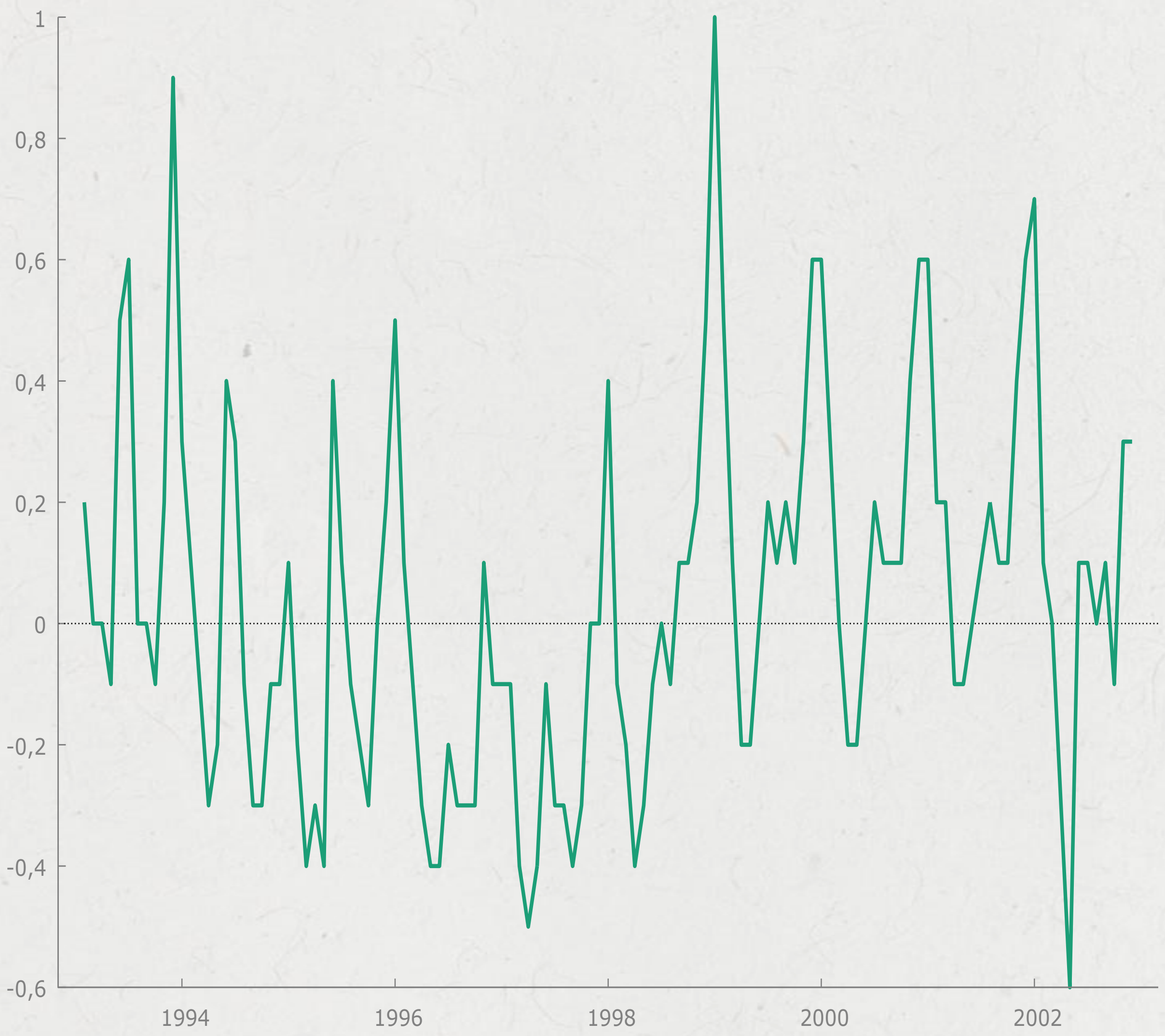

Фактические значения курса акций Сбербанка
400
350
300
250
200
150
100
50
0
2008-01-01
2010-01-01
2012-01-01
2014-01-01
2016-01-01
2018-01-01
2020-01-01
2022-01-01


Первые конечные разности
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
2008-01-01
2010-01-01
2012-01-01
2014-01-01
2016-01-01
2018-01-01
2020-01-01
2022-01-01

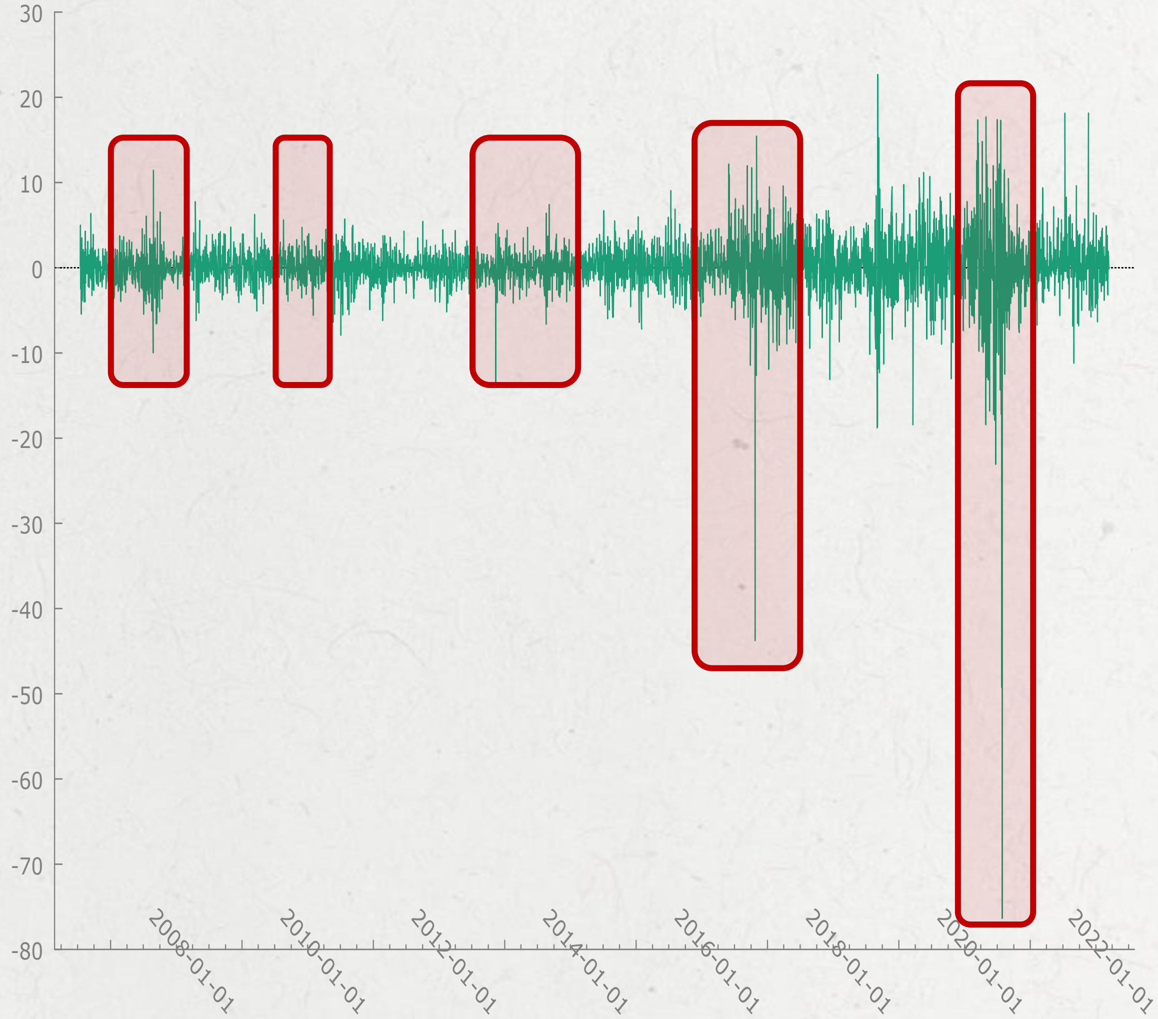

Фактические значения курса
акций Сбербанка
400
350
300
250
200
150
100
50
0
2008-01-01
2010-01-01
2012-01-01
2014-01-01
2016-01-01
2018-01-01
2020-01-01
2022-01-01


Первые конечные разности
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
2008-01-01
2010-01-01
2012-01-01
2014-01-01
2016-01-01
2018-01-01
2020-01-01
2022-01-01

«Белый шум» – случайные колебания



«Кластеры» волатильности



Зачем учитывать условную гетероскедастичность при моделировании?

Эффективность. Мы знаем, что корректный учет гетероскедастичности позволяет получить более точные оценки коэффициентов.

Предсказание волатильности. Иногда при анализе динамики финансовых переменных важно уметь предсказывать их волатильность (дисперсию). Например, для оценки степени рискованности актива.

Ключевой особенностью моделей *ARCH* (AutoRegressive Conditional Heteroscedasticity) и *GARCH* (*General ARCH*) является разграничение условных и безусловных дисперсий.

В то время как безусловная матрица ковариаций для представляющих интерес переменных может быть неизменной во времени, условные дисперсии и ковариации часто зависят нетривиальным образом от состояний изучаемых явлений или процессов в прошлом.

Условной дисперсией называется дисперсия случайной переменной, обусловленная информацией о других случайных переменных, то есть дисперсия, найденная при условии наличия знаний о дисперсии в предыдущие моменты времени: $\sigma_t^2 = D(u_t \mid u_{t-1}, u_{t-2\ldots})$

ARCH(1) - модель первого порядка имеет вид: $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}$

где $u_t = y_t - y_t$ - остатки полученные после предварительной оценки какой-либо модели временного ряда (например, AR)

Модель ARCH(p) (порядок авторегрессии дисперсии – p) имеет вид:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_p u_{t-p}^2$$