

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. Ломоносова  
ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ  
КАФЕДРА ММАЭ

КУРСОВАЯ РАБОТА  
**«ТЕСТЫ ВРЕМЕННЫХ РЯДОВ НА СТАЦИОНАРНОСТЬ»**

Выполнил студент  
группы Э202:  
Жижерин С.Н.  
Научный руководитель:  
доцент Артамонов Д.В.

Москва  
2020

# Оглавление

1	Введение .....	3
2	Теоретическая часть .....	4
2.1	Основные понятия .....	4
2.2	Тесты на стационарность .....	6
2.2.1	О тестах Дики – Фуллера .....	6
2.2.2	Простой $\rho$ -тест Дики – Фуллера .....	7
2.2.3	Простой $t$ -тест Дики – Фуллера .....	13
3	Расчётная часть .....	15
3.1	Введение в расчётную часть .....	15
3.2	Расчёты для $\rho$ -статистики .....	16
3.3	Расчёты для $t$ -статистики .....	21
3.4	Пример теста на стационарность .....	23
4	Заключение .....	24
5	Литература .....	25
6	Приложение: программа на Python 3.7 .....	26

# 1 Введение

Предметом этой работы являются тесты временных рядов на стационарность. Временной ряд в эконометрике рассматривается как реализация некоторого случайного процесса. В литературе используется сокращение DGP от английского “Data Generating Process”. Такой подход весьма интересен и полезен, поскольку многим экономическим явлениям присуща случайность, понимаемая как принципиально неустранимая неопределённость. Простой иллюстрацией этого тезиса могут служить финансовые рынки: если устранить неопределённость на таком рынке, он, очевидно, перестанет функционировать. Тем самым, часто следует рассматривать многократно меняющие свои значения с течением времени экономические данные как реализации случайного процесса. Таким образом, возникает задача об отыскании DGP, то есть о подборе подходящей модели случайного процесса к реальным данным. Существуют целые классы моделей, предполагающие наличие у процесса различных свойств. Заметим, что одной из главных целей такого подхода является прогнозирование будущих значений ряда данных на основе подобранной модели случайного процесса. Выясняется, что для успешного прогнозирования нужно, чтобы DGP обладал некоторыми полезными свойствами. Одним из важных свойств является стационарность. Стационарность можно описать как независимость ряда важных свойств случайного процесса от времени. Понятно, что в контексте прогнозирования такое свойство выглядит привлекательно. Более строгое определение различных типов стационарности даётся в теоретической части работы. Для того, чтобы иметь возможность с определённой точностью отнести процесс, соответствующий некоторому временному ряду к классу стационарных или нестационарных, придуманы различные тесты. Целью данной работы является обзор теста Дики – Фуллера (Dickey – Fuller) с выводом асимптотических распределений его тестовых статистик. Также, в расчётной части работы рассматривается вопрос об отыскании плотностей соответствующих распределений и их критических значений при помощи метода Монте-Карло.

Главной теоретической основой этой работы является учебник Фумио Хаяши “Эконометрика”. Все утверждения, которые формулируются в теоретической части, но не доказываются, являются, по существу, ссылками на эту книгу. Также для связности и ясности изложения основная терминология построена по образцу терминологии “Эконометрики”. Названия некоторых терминов дублируются на английском, поскольку это позволяет избежать неточностей.

## 2 Теоретическая часть

### 2.1 Основные понятия

Рассматривается последовательность случайных величин  $\{X_t\}$ , называемая также случайным процессом. Прежде всего, следует определить понятие стационарности случайного процесса. Это свойство предполагает независимость некоторых свойств процесса от времени. Существует три набора условий различной силы, позволяющих гарантировать полезные свойства процесса. Наиболее слабые требования таковы:

**Определение** (covariance-stationary process, слабостационарный процесс)

Случайный процесс  $\{X_t\}$  называется стационарным в широком смысле, если:

- 1)  $\forall t, \exists E(X_t) = \mu$  не зависящее от  $t$ ;
- 2)  $\forall k \in \mathbb{Z}, \gamma(k) = \text{cov}(X_t, X_{t+k}) < \infty$  и зависит только от  $k$ .

Заметим, что из данного определения непосредственно следует нечётность автоковариационной функции ( $\gamma(k) \equiv \gamma_k$ ) стационарного в широком смысле процесса:  $\gamma_k = \gamma_{-k}$ . Сформулируем более сильное условие:

**Определение** (strictly stationary process)

Случайный процесс  $\{X_t\}$  называется стационарным в узком смысле, если совместное распределение произвольного набора случайных величин  $(X_i, X_{i_1}, \dots, X_{i_r})$  не зависит от  $i$ .

Как видно из определений, стационарность в узком смысле влечёт за собой стационарность в широком смысле при условии конечности ковариаций и дисперсий. К требованию стационарности в узком смысле можно добавить ещё одно условие, позволяющее использовать теорему (2), которая будет сформулирована в дальнейшем.

**Определение** (ergodic stationary process)

Стационарный в узком смысле случайный процесс  $\{X_t\}$  называется эргодическим, если справедливо:

$$\forall f, g (f: \mathbb{R}^k \rightarrow \mathbb{R}, g: \mathbb{R}^l \rightarrow \mathbb{R}), \lim_{n \rightarrow \infty} |E(f(X_i, \dots, X_{i+k}))g(X_{i+n}, \dots, X_{i+n+l})| = |E(f(X_i, \dots, X_{i+k}))| |E(g(X_{i+n}, \dots, X_{i+n+l}))|, \text{ где } f \text{ и } g - \text{ограниченные функции.}$$

Тем самым, две достаточно далеко друг от друга расположенные случайные величины эргодического процесса распределены практически независимо.

Для удобства дальнейших выкладок перейдём к рассмотрению временного ряда  $\{y_t\}$ , который будем понимать как реализацию некоторого случайного процесса. Тем самым,  $y_t$  – случайная величина. Такие обозначения используются, чтобы подчеркнуть, что предполагается наличие некоторого набора наблюдений.

### **Определение (интегрированные процессы)**

Интегрированным порядка нуль  $I(0)$  процессом называется стационарный в узком смысле случайный процесс с положительной долгосрочной дисперсией, где долгосрочная дисперсия стационарного процесса  $\{y_t\}$  определяется как  $\lim_{n \rightarrow \infty} \text{Var}(\sqrt{n}\bar{y})$ , если этот предел существует. Тем самым,  $I(0)$  процесс представим в виде  $\delta + U_t$ , где  $\{U_t\}$  – строго стационарный процесс с положительной долгосрочной дисперсией. Интегрированным порядка  $d$   $I(d)$  процессом называется  $\{y_t\}$ , для которого  $\Delta^d y_t$  является  $I(0)$ . Оператор разности определяется как  $\Delta^d = (1 - L)^d$ , где  $Ly_t = y_{t-1}$  – оператор запаздывания.

Для того, чтобы в дальнейшем получить асимптотические распределения тестовых статистик, введём дополнительное ограничение на рассматриваемые  $I(0)$  процессы:

### **Определение**

Процесс называется линейным  $I(0)$ , если он представим в виде  $\delta + U_t$ , где  $\{U_t\}$  – линейный:  $U_t = \psi(L)\varepsilon_t$ ,  $\psi(L) \equiv \psi_0 + \psi_1 L + \psi_2 L^2 + \dots$ ,  $t \in \mathbb{Z}$

Притом,  $\{\varepsilon_t\}$  – процесс независимого белого шума,  $E(\varepsilon_t) \equiv \sigma^2 > 0$  и верно:  $\sum_{j=0}^{\infty} j|\psi_j| < \infty$  и  $\psi(L) \neq 0$ .

В ходе дальнейшего изложения под  $I(0)$  понимаются именно такие процессы.

Таким образом, введены основные понятия, нужные для того, чтобы перейти непосредственно к содержанию теоретической части работы.

## 2.2 Тесты на стационарность

### 2.2.1 О тестах Дики – Фуллера

В этой работе рассматривается тест Дики – Фуллера, позволяющий проверять нулевую гипотезу о том, что процесс является случайным блужданием. Случайное блуждание – частный случай процесса  $I(1)$ . Такой процесс не стационарен, однако разность его уровней обладает свойством стационарности, поскольку является  $I(0)$ . Процессы  $I(1)$  также называют процессами единичного корня, поскольку их особенностью является наличие у лагового полинома корня модуля единица.

Тест назван в честь своих авторов, американских статистиков Д.Дики (David Dickey) и У.Фуллера (Wayne Fuller). Сам тест и его усовершенствованная версия активно используются при анализе временных рядов. Это обусловлено важностью стационарности и распространённостью процессов авторегрессии  $AR(p)$  различных порядков, для которых и применим данный тест. К преимуществам теста можно отнести также его простоту: требуется лишь рассчитать по несложной формуле значение соответствующего выражения, называемого тестовой статистикой и сопоставить его с табличным, чтобы иметь возможность делать выводы о стационарности. В этой работе рассматриваются две наиболее простые версии теста Дики – Фуллера: простой  $\rho$ -тест и простой  $t$ -тест. Они используются в тех случаях, когда процесс предположительно описывается моделью авторегрессии первого порядка с нулевым математическим ожиданием.

Применение тестов основано на том, что у тестовых статистик, являющихся функциями от выборки и её объёма есть асимптотические распределения при нулевой гипотезе. Тем самым, зная такое распределение и рассчитанное по конкретным данным значение тестовой статистики, можно утверждать, что нулевая гипотеза верна с некоторой вероятностью. Вопросу об отыскании такой вероятности для конкретных объёмов выборки и значений статистики посвящена расчётная часть этой работы.

## 2.2.2 Простой $\rho$ -тест Дики – Фуллера

Рассмотрим модель процесса авторегрессии первого порядка  $AR(1)$ :  $y_t = \rho y_{t-1} + \varepsilon_t$ , где  $\{\varepsilon_t\}$  – независимый белый шум.

Предполагая заранее, что процесс является либо  $I(0)$ , либо случайным блужданием без сдвига, получаем ограничение:  $-1 < \rho \leq 1$ . Таким образом, нулевая гипотеза принимает вид  $\rho = 1$ , а альтернативная гипотеза:  $\rho < 1$ . Тест основан на точечной оценке параметра  $\rho$  по методу наименьших квадратов. Пусть имеется  $T + 1$  наблюдение  $(y_0, y_1, \dots, y_T)$ . Относительно начального условия предположим, что  $E(y_0^2) < \infty$ .

$$\begin{aligned} S(\rho) &= \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \rho y_{t-1})^2 \rightarrow \min \\ S'(\rho) &= -2 \sum_{t=1}^T y_{t-1} (y_t - \rho y_{t-1}) = 0 \\ \sum_{t=1}^T y_{t-1} y_t &= \rho \sum_{t=1}^T y_{t-1}^2 \\ \hat{\rho} &= \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} \end{aligned}$$

Необходимое условие экстремума является и достаточным, так как  $S(\rho)$  – выпуклая вниз функция. Таким образом, получена МНК-оценка параметра. При нулевой гипотезе имеем ошибку выборки:

$$\hat{\rho} - 1 = \frac{\sum_{t=1}^T y_{t-1} \Delta y_t}{\sum_{t=1}^T y_{t-1}^2}$$

Выясняется, что у случайной величины  $T(\hat{\rho} - 1)$  есть асимптотическое распределение. Для вывода этого распределения необходимо сформулировать и обосновать ряд утверждений. Начнём с основной теоремы.

### Теорема (1)

Пусть  $\{y_t\}$  – процесс  $I(1)$  без сдвига такой, что  $E(y_0^2) < \infty$  и  $\{\Delta y_t\}$  удовлетворяет сформулированным ранее условиям для  $I(0)$ . Обозначим дисперсию  $\{\Delta y_t\}$  за  $\gamma_0$ , долгосрочную дисперсию  $\{\Delta y_t\}$  за  $\lambda^2$ . Тогда:

$$(a) \frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2 \xrightarrow{d} \lambda^2 \int_0^1 W(r)^2 dr$$

$$(b) \frac{1}{T} \sum_{t=1}^T y_{t-1} \Delta y_t \xrightarrow{d} \frac{\lambda^2}{2} W(1)^2 - \frac{\gamma_0}{2}$$

Притом эти случайные величины сходятся как случайный вектор.

$W(\cdot)$  здесь – винеровский процесс, являющийся непрерывным по времени  $t \in [0, 1]$  аналогом случайного блуждания. Он обладает следующими свойствами:

- 1)  $W(0) = 0$ ;
- 2)  $\forall t_1, \dots, t_k (0 \leq t_1 < \dots < t_k \leq 1)$ ,  $W(t_2) - W(t_1), \dots, W(t_k) - W(t_{k-1})$  – независимые нормально распределённые случайные величины:  
 $W(s) - W(t) \sim N(0, (s - t))$ ;
- 3) Для всякой реализации процесса,  $W(t)$  непрерывен по  $t$  с вероятностью 1.

Обоснуем пункт (b) этого утверждения. Чтобы это сделать, нужно сформулировать три теоремы. Первая из них уже была упомянута при введении понятия эргодического процесса.

### Теорема (2) (Ergodic theorem)

Пусть  $\{z_i\}$  – стационарный эргодический процесс с математическим ожиданием  $\mu$ . Тогда  $\bar{z}_n \equiv \frac{1}{n} \sum_{i=1}^n z_i \xrightarrow[a.s.]{a.s.} \mu$ , где  $\xrightarrow[a.s.]{} a.s.$  означает сходимость почти наверное, из которой следует сходимость по вероятности.

Заметим, что из сходимости по вероятности в свою очередь следует сходимость по распределению:  $\{z_n\} \xrightarrow{p} z \implies \{z_n\} \xrightarrow{d} z$ .

### Теорема (3) (ЦПТ для процессов $MA(\infty)$ )

Пусть  $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ , где  $\{\varepsilon_t\}$  – независимый белый шум и  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . Тогда:

$$\sqrt{n}(\bar{y} - \mu) \xrightarrow{d} N(0, \sum_{j=-\infty}^{+\infty} \gamma_j)$$



Дисперсия такого распределения равна долгосрочной дисперсии  $\{y_t\}$ . Обоснуем этот факт.

### Утверждение

Пусть  $\{y_t\}$  – слабостационарный процесс,  $\gamma_j$  – его автоковариации и  $\sum_{j=0}^{\infty} \gamma_j < \infty$ . Тогда

$$\lambda^2 = \lim_{n \rightarrow \infty} \text{Var}(\sqrt{n}\bar{y}) = \sum_{j=-\infty}^{+\infty} \gamma_j.$$

### Доказательство

Воспользуемся формулой  $\text{Var}(\xi + \eta) = \text{Var}(\xi) + 2\text{cov}(\xi, \eta) + \text{Var}(\eta)$ :

$$\begin{aligned} n\text{Var}(\sqrt{n}\bar{y}) &= n^2\text{Var}(\bar{y}) = \text{Var}\left(\sum_{i=1}^n y_i\right) \\ &= \text{Var}(y_1) + 2\text{cov}\left(y_1, \sum_{i=2}^n y_i\right) + \text{Var}\left(\sum_{i=2}^n y_i\right) \\ &= \sum_{i=1}^n \text{Var}(y_i) + 2\left(\text{cov}\left(y_1, \sum_{i=2}^n y_i\right) + \text{cov}\left(y_2, \sum_{i=3}^n y_i\right) + \dots + \text{cov}(y_{n-1}, y_n)\right) \\ &= n\gamma_0 + 2\left(\sum_{i=1}^{n-1} \gamma_i + \sum_{i=1}^{n-2} \gamma_i + \dots + \gamma_1\right) \\ &= n\gamma_0 + 2((n-1)\gamma_1 + (n-2)\gamma_2 + \dots + \gamma_{n-1}) \\ &= n\gamma_0 + 2\sum_{j=1}^{n-1} (n-j)\gamma_j \Rightarrow \\ &\Rightarrow \text{Var}(\sqrt{n}\bar{y}) = \gamma_0 + 2\sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right)\gamma_j \\ &= \gamma_0 + 2\sum_{j=1}^{n-1} \gamma_j - \frac{2}{n}\sum_{j=1}^{n-1} j\gamma_j \end{aligned}$$

Покажем теперь, что предел третьего слагаемого равен нулю:

$$\frac{2}{n}\sum_{j=1}^{n-1} j\gamma_j = \frac{2}{n}(\gamma_1 + 2\gamma_2 + \dots + (n-1)\gamma_{n-1}) = \frac{2}{n}\sum_{i=1}^{n-1}\sum_{j=i}^{n-1} \gamma_j$$

По условию  $\sum_{j=0}^{\infty} \gamma_j < \infty$ , тогда по критерию Коши сходимости ряда:

$$\forall \varepsilon > 0 \exists n_0: \forall n \geq n_0 \forall p \in \mathbb{N}, \left| \sum_{j=n+1}^{n+p} \gamma_j \right| < \varepsilon$$

Тогда двойную сумму можно представить в виде:

$$\frac{2}{n} \sum_{i=1}^{n_0} \sum_{j=i}^{n-1} \gamma_j + \frac{2}{n} \sum_{i=n_0+1}^{n-1} \sum_{j=i}^{n-1} \gamma_j$$

Предел первого слагаемого равен нулю. Это так, поскольку, устремив  $n$  к бесконечности, мы получим сумму конечного числа ( $n_0$ ) стремящихся к нулю выражений вида: удвоенная сумма сходящегося ряда, делённая на  $n$ . Второе слагаемое также стремится к нулю при  $n \rightarrow \infty$ :

$$\left| \frac{1}{n} \sum_{i=n_0+1}^{n-1} \sum_{j=i}^{n-1} \gamma_j \right| = \frac{1}{n} \sum_{i=n_0+1}^{n-1} \left| \sum_{j=i}^{n-1} \gamma_j \right| < \frac{1}{n} \sum_{i=n_0+1}^n |\varepsilon| \leq \frac{n\varepsilon}{n} = \varepsilon$$

Таким образом:

$$\forall \varepsilon > 0 \exists n_1 = n_0 + 3 : \forall n \geq n_1, \left| \frac{1}{n} \sum_{i=n_0+1}^{n-1} \sum_{j=i}^{n-1} \gamma_j - 0 \right| < \varepsilon$$

То есть для второго слагаемого выполнено определение бесконечно малой последовательности. Тем самым, для  $\lambda^2$  имеем:

$$\lambda^2 = \lim_{n \rightarrow \infty} \left( \gamma_0 + 2 \sum_{j=1}^{n-1} \gamma_j \right) = \sum_{j=-\infty}^{+\infty} \gamma_j$$

Последнее равенство справедливо, поскольку для слабостационарного процесса  $\gamma_j = \gamma_{-j}$ . ■

Теперь приступим непосредственно к доказательству пункта (b) теоремы (1). Пункт (a) при этом доказывать не будем, поскольку для этого потребовалось бы дополнительно сформулировать более сложную предельную теорему. Упростим выражение:

$$\begin{aligned}
y_{t-1}\Delta y_t &= y_{t-1}y_t - y_{t-1}^2 = y_t(y_t - \Delta y_t) - y_{t-1}^2 = y_t^2 - y_t\Delta y_t - y_{t-1}^2 \\
&= y_t^2 - y_{t-1}^2 - \Delta y_t(y_t + y_{t-1}) = y_t^2 - y_{t-1}^2 - \Delta y_t^2 - y_{t-1}\Delta y_t \Rightarrow \\
&\Rightarrow y_{t-1}\Delta y_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \Delta y_t^2) \Rightarrow \\
&\Rightarrow \sum_{t=1}^T y_{t-1}\Delta y_t = \frac{1}{2}\sum_{t=1}^T (y_t^2 - y_{t-1}^2) - \frac{1}{2}\sum_{t=1}^T \Delta y_t^2 \\
&= \frac{1}{2}(y_T^2 - y_0^2) - \frac{1}{2}\sum_{t=1}^T \Delta y_t^2 \Rightarrow \\
&\Rightarrow \frac{1}{T}\sum_{t=1}^T y_{t-1}\Delta y_t = \frac{1}{2}\left(\frac{y_T}{\sqrt{T}}\right)^2 - \frac{1}{2}\left(\frac{y_0}{\sqrt{T}}\right)^2 - \frac{1}{2T}\sum_{t=1}^T \Delta y_t^2
\end{aligned}$$

Рассмотрим третье слагаемое. Процесс  $\{\Delta y_t\}$  эргодический. Тогда из определения следует, что  $\{\Delta y_t^2\}$  – тоже эргодический. Используя теорему (2), получаем, что третье слагаемое сходится по распределению к  $-\frac{\gamma_0}{2}$ , где  $\gamma_0 = E(\Delta y_t^2) = Var(\Delta y_t)$ .

Теперь рассмотрим второе слагаемое. К нему применимо известное неравенство Чебышёва:  $P(|\xi - M\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$

Имеем:  $Var\left(\frac{y_0}{\sqrt{T}}\right) = \frac{Var(y_0)}{T} = \frac{E(y_0^2)}{T}$  Тем самым, второе слагаемое сходится по вероятности, а значит и по распределению к своему математическому ожиданию, то есть к нулю. Справедливость этих выводов при возведении в квадрат гарантируется теоремой Слуцкого, которая утверждает, что непрерывное преобразование не нарушает сходимости.

Наконец, рассмотрим первое слагаемое:

$$\frac{y_T}{\sqrt{T}} = \frac{1}{\sqrt{T}}\left(\sum_{t=1}^T \Delta y_t + y_0\right) = \sqrt{T}\Delta y_t + \frac{y_0}{\sqrt{T}}$$

Второе слагаемое полученного выражения сходится по вероятности к нулю. Из определения линейного процесса  $I(0)$  следует, что  $\{\Delta y_t\}$  удовлетворяет условиям теоремы (3). Тогда в силу  $W(1) \sim N(0, 1)$  имеем:

$$\frac{y_T}{\sqrt{T}} \xrightarrow{d} \lambda W(1)$$

Таким образом, пункт (b) теоремы (1) доказан. ■

Вернёмся теперь к вопросу об отыскании асимптотического распределения  $\rho$ -статистики. Из сформулированных утверждений следует:

$$T(\hat{\rho} - 1) = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} \Delta y_t}{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2} \xrightarrow{d} \frac{\frac{\lambda^2}{2} W(1)^2 - \frac{\gamma_0}{2}}{\lambda^2 \int_0^1 W(r)^2 dr}$$

Поскольку рассматривается процесс, удовлетворяющий условиям теоремы (1), числитель и знаменатель дроби сходятся по распределению как вектор к вектору из соответствующих случайных величин. Тогда по теореме Slutsky, поскольку  $T(\hat{\rho} - 1)$  – непрерывная функция числителя и знаменателя, она сходится по распределению к соответствующей дроби.

Выражение можно упростить, так как  $\{\Delta y_t\}$  – процесс независимого белого шума и, следовательно,  $\lambda^2 = \gamma_0$  в силу полученного для  $\lambda^2$  выражения. Тем самым, окончательно имеем:

$$T(\hat{\rho} - 1) \xrightarrow{d} \frac{\frac{1}{2} (W(1)^2 - 1)}{\int_0^1 W(r)^2 dr} \equiv DF_\rho$$

Полученная тестовая статистика называется  $\rho$ -статистикой Дики – Фуллера и используется для тестирования гипотезы о том, что процесс является случайным блужданием без сдвига. Поскольку нулевая гипотеза не накладывает ограничений ни на величину дисперсии  $\{\varepsilon_t\}$ , ни на значение начального условия  $y_0$ , статистика  $DF_\rho$  и её критические значения не зависят от этих параметров.

### 2.2.3 Простой $t$ -тест Дики – Фуллера

При тех же предпосылках можно получить распределение  $t$ -статистики. Запишем и преобразуем выражение для обычной  $t$ -статистики при нулевой гипотезе  $\rho = 1$ :

$$t = \frac{\hat{\rho} - 1}{s \div \sqrt{\sum_{t=1}^T y_{t-1}^2}} = \frac{\sum_{t=1}^T y_{t-1} \Delta y_t}{s \sqrt{\sum_{t=1}^T y_{t-1}^2}} = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} \Delta y_t}{s \sqrt{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}}$$

Здесь  $s$  – стандартная ошибка, определяемая следующим образом:

$$s \equiv \sqrt{\frac{1}{T-1} \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2}$$

Для последующего вывода асимптотического распределения  $t$ -статистики покажем, что  $s^2$  сходится по вероятности, а значит и по распределению к величине  $\gamma_0 = E(\Delta y_t^2) = \text{Var}(\Delta y_t)$ :

$$\begin{aligned} s^2 &= \frac{1}{T-1} \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2 \\ &= \frac{1}{T-1} \sum_{t=1}^T (y_t^2 - 2\hat{\rho} y_t y_{t-1} + \hat{\rho}^2 y_{t-1}^2) \\ &= \frac{1}{T-1} \left( \sum_{t=1}^T (y_{t-1}^2 + 2\Delta y_t y_{t-1} + \Delta y_t^2 - 2\hat{\rho} y_{t-1}^2 - 2\hat{\rho} \Delta y_t y_{t-1} + \hat{\rho}^2 y_{t-1}^2) \right) \\ &= \frac{1}{T-1} \left( (\hat{\rho} - 1)^2 \sum_{t=1}^T y_{t-1}^2 - 2(\hat{\rho} - 1) \sum_{t=1}^T \Delta y_t y_{t-1} + \sum_{t=1}^T \Delta y_t^2 \right) \\ &= \frac{1}{T-1} (T(\hat{\rho} - 1))^2 \frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2 - \frac{2}{T-1} (T(\hat{\rho} - 1)) \frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1} + \\ &\quad + \frac{T}{T-1} \frac{1}{T} \sum_{t=1}^T \Delta y_t^2 \end{aligned}$$

Первое и второе слагаемые в полученном выражении устроены схоже. В каждом из них первый множитель – общий член сходящейся к нулю последовательности, второй множитель – функция от  $\rho$ -статистики, которая сходится по распределению к  $DF_\rho$ , третий множитель – выражение из пунктов (а) и (b) теоремы (1) соответственно. Тем самым, первое и второе слагаемые сходятся по вероятности к нулю. В третьем слагаемом  $\frac{T}{T-1} \rightarrow 1$ , а сходимость по вероятности второго множителя к  $\gamma_0$  следует из теоремы (2). Рассуждения для точно такого же выражения уже приводились выше при доказательстве пункта (b) теоремы (1). Таким образом,  $s^2 \xrightarrow{d} \gamma_0$ . ■

Теперь пользуясь этим фактом, теоремой (1), теоремой Слуцкого и тем, что при сделанных предположениях  $\lambda^2 = \gamma_0$ , получаем:

$$t = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} \Delta y_t}{s \sqrt{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}} \xrightarrow{d} \frac{\frac{1}{2} (W(1)^2 - 1)}{\sqrt{\int_0^1 W(r)^2 dr}} \equiv DF_t$$

Заметим, что, как и в случае с  $\rho$ -статистикой, критические значения  $t$ -статистики не зависят ни от величины дисперсии процесса белого шума, ни от начального условия.

Помимо прочего, можно видеть, что для выражений  $DF_\rho$  и  $DF_t$  числитель и знаменатель содержат один и тот же случайный процесс и, следовательно, зависимы. Это обстоятельство нужно учитывать при компьютерном моделировании случайных величин, о котором речь пойдёт в расчётной части работы.

### 3 Расчётная часть

#### 3.1 Введение в расчётную часть

В теоретической части работы тезисно изложен вывод асимптотических распределений тестовых статистик для простого теста Дики – Фуллера. Возникает вопрос о практическом применении этого теоретического результата при анализе временных рядов. Отметим, что гораздо чаще на практике применяется усовершенствованный тест, однако в расчётной части работы будут рассмотрены проблемы, возникающие при применении простого теста. Это позволяет упростить выкладки и соблюсти последовательность рассуждений.

Сходимость по распределению выборочных значений тестовых статистик действительно является важным результатом, прежде всего, по двум причинам. Первая из них – сам факт сходимости выборочных статистик. Этот факт позволяет при помощи моделирования методом Монте-Карло получать с некоторой степенью точности критические значения тестовых статистик для конкретных объёмов выборки или, что более удобно, получать  $p$ -значение для данной выборки. Вторая причина – конкретное выражение для двух случайных величин, к которым сходятся по распределению тестовые статистики. Такое выражение позволяет с некоторой степенью точности моделировать реализации этих случайных величин, получая теоретическую функцию распределения. Отметим, однако, что такое моделирование и оценка его погрешности является несколько менее тривиальной задачей. Это связано с тем, что в выражения для  $DF_\rho$  и  $DF_t$  входит стохастический интеграл  $\int_0^1 W(r)^2 dr$ , который трудно вычислить в явном виде, где под “явным видом” подразумевается представление в виде комбинации нескольких независимых случайных величин, притом независимых не только друг от друга, но и от случайной величины в числителе  $DF_\rho$  и  $DF_t$ ,  $W(1)^2$ . Подобное представление было бы весьма удобно для компьютерного моделирования.

Впрочем, реальные временные ряды имеют конечную длину и потому теоретическая функция распределения играет меньшую роль при анализе реальных данных. Однако, приближённое моделирование  $DF_\rho$  и  $DF_t$  позволяет понять вид графика теоретической плотности распределения.

Приводимые расчёты и графики авторские, использован язык программирования Python 3.7, полный текст программы вынесен в приложение к работе.

### 3.2 Расчёты для р-статистики

Для лучшего понимания теории произведём ряд расчётов. Программа, представленная в приложении к работе, имеет достаточно простую структуру. Все её элементы представляют собой функции, что позволяет многократно воспроизводить вычисления, передавая функциям соответствующие аргументы. Основной принцип построения этих функций – использование векторных операций вместо циклов для ускорения вычислений. В тексте расчётной части приводятся только вызовы функций (кроме функций и частей программы, отвечающих за визуализацию, которые не приводятся ни в тесте, ни в приложении из-за громоздкости).

Выясним сначала вид теоретической плотности для случайной величины  $DF_\rho$ :

$$DF_\rho \equiv \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr}$$

Из известной в математической статистике теоремы Колмогорова о сходимости эмпирической функции распределения к теоретической, в частности, следует, что скорость этой сходимости имеет порядок  $\frac{1}{\sqrt{n}}$ . Эта теорема будет использоваться в качестве оценки погрешности при вычислении критических точек распределений. Сформулируем её более строго:

**Теорема (4)** (Теорема Колмогорова)

$$P(\sqrt{n}D_n < x) \rightarrow Q(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}, n \rightarrow \infty;$$

где  $D_n = \sup(|F_n(x) - F(x)|)$

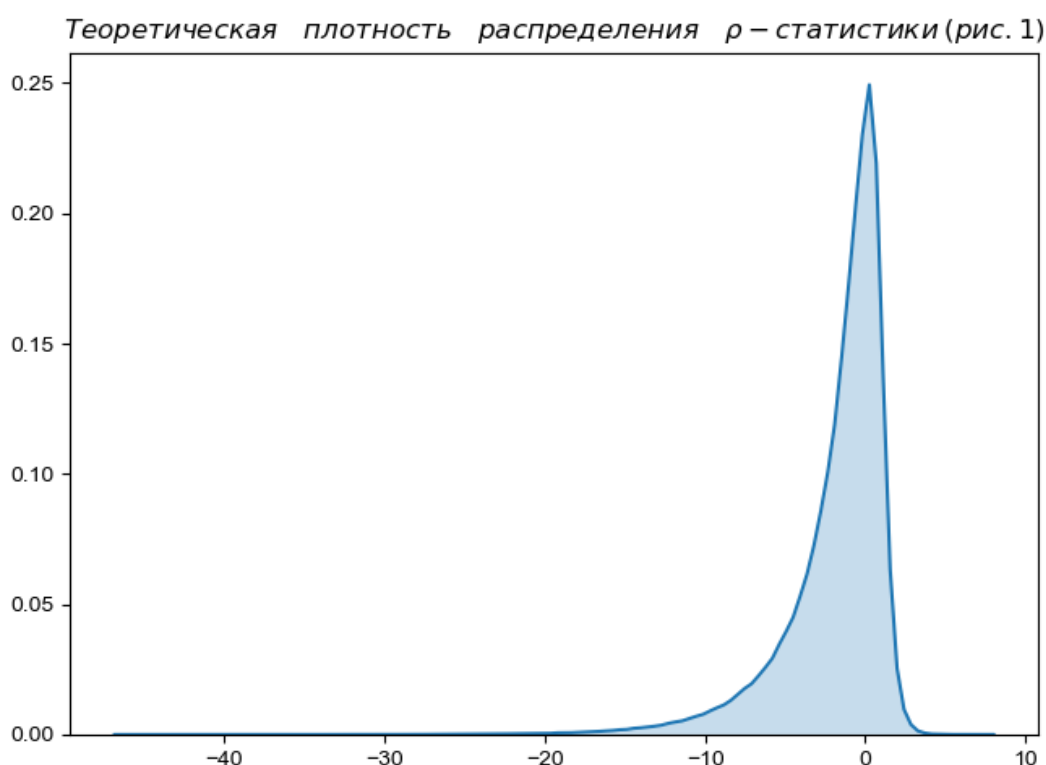
По таблице критических точек распределения Колмогорова находим  $P(\sqrt{n}D_n \geq 1,63) = 0,01$ . Следовательно,  $P\left(\sup(|F_n(x) - F(x)|) < \frac{1,63}{\sqrt{n}}\right) = 0,99$ . Это выражение даёт оценку погрешности для эмпирической функции распределения, вычисленной по  $n$  реализациям случайной величины: с вероятностью 0,99 эта погрешность меньше величины  $\frac{1,63}{\sqrt{n}}$ .



В случае с  $DF_\rho$  помимо этой основной погрешности возникает ещё и погрешность вычисления интеграла от квадрата винеровской траектории. Каждая реализация этого интеграла вычисляется путём генерирования  $m$  независимых хи-квадрат распределённых случайных величин, каждая из которых является квадратом накопленной суммы нормально распределённых величин:

$$\xi_k = (\gamma_1 + \dots + \gamma_k)^2, \quad \gamma_i \sim N\left(0, \frac{1}{m}\right)$$

Ясно, что  $m$  таких величин образуют некоторую траекторию, являющуюся графиком кусочно-линейной функции, приближающей процесс  $W(r)^2$  на отрезке  $[0; 1]$ . Метод средних прямоугольников позволяет точно проинтегрировать такую функцию. Грубая оценка погрешности такого интегрирования позволяет сделать вывод о том, что можно положить  $m = 1000$  без существенного ущерба для наглядности. Опять же, для наглядности воспользуемся вместо гистограммы ядерной оценкой плотности. Чтобы сгенерировать массив из  $10^6$  реализаций  $DF_\rho$  с  $m = 5000$  произведём такой вызов функции: `calculate_theoretical(rho_theoretical(1000000, 5000))`. Построим график:



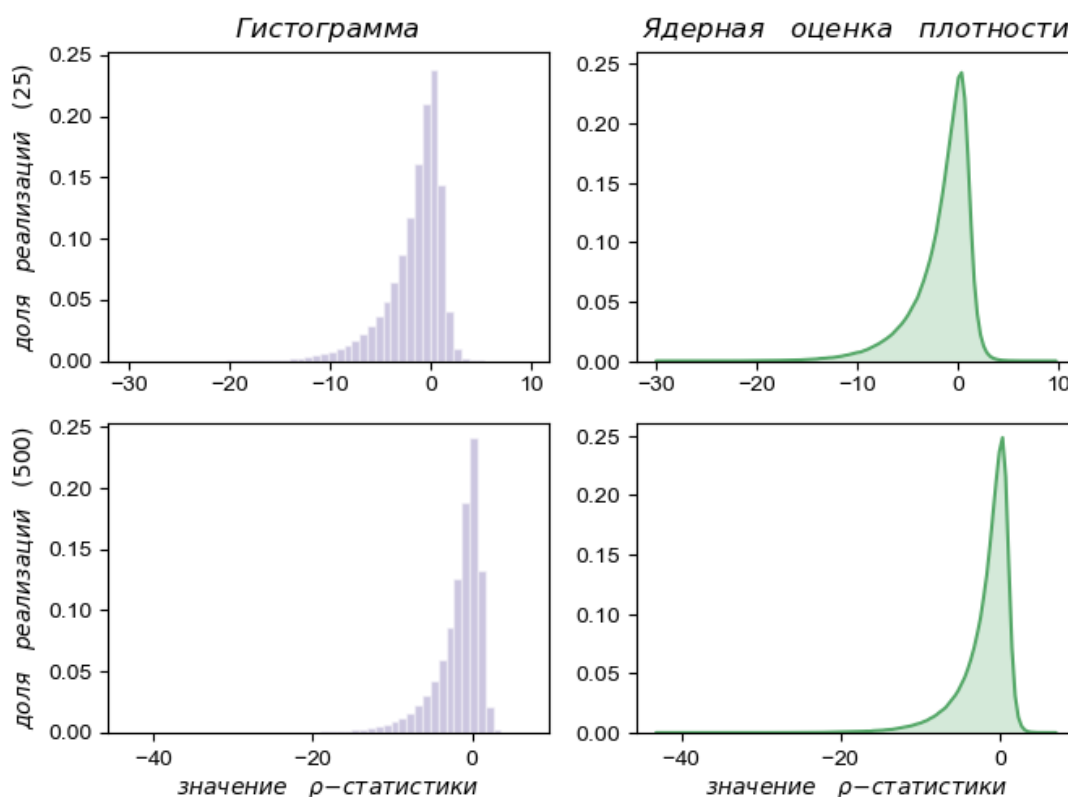
При больших объёмах выборки плотность распределения должна иметь схожий вид.

Перейдём теперь к визуализации плотности для конечных объёмов выборки. Здесь воспользуемся самым фактом сходимости по распределению тестовой статистики. Будем многократно генерировать случайное блуждание и рассчитывать по нему статистику:

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

Несмотря на использование векторных операций, многократное вычисление этого выражения весьма трудоёмко. Приведём графики плотности распределения, построенные на основе  $10^6$  случайных блужданий длины 25 и 500. Эти числа определяют приблизительные границы применимости такого способа отыскания плотности: если длина ряда меньше 20, этого может быть недостаточно для получения распределения, при длине ряда большей 500, распределение сближается с теоретическим, для которого расчёты проще при условии точной оценки погрешности.

*Плотности распределения для рядов длины 25 и 500 (рис. 2)*



Как и предполагалось, для большой длины ряда (500), плотность сближается с теоретической, которая соответствует бесконечному ряду.

Приведённые выше графики дают представление о виде плотности. Однако, для проведения тестов на данных нужны именно критические точки распределения. Под критической точкой распределения понимается некоторая граница, для которой вероятность того, что значение случайной величины окажется за её пределами, мала. Как уже отмечалось в теоретической части, при введённых предпосылках  $-1 < \rho \leq 1$ . Притом, все распределения получены при нулевой гипотезе  $\rho = 1$ . Из вида тестовой статистики  $T(\hat{\rho} - 1)$  и графиков ясно, что в данном случае критическую точку  $x_\alpha$  уместно определить как величину, для которой верно  $P(T(\hat{\rho} - 1) < x_\alpha) = \alpha$ . Вероятность  $\alpha$  называется уровнем значимости и выбирается тем меньшей, чем большую надёжность выводов нужно получить на основании теста, если нулевая гипотеза отвергается и наоборот. Иначе говоря, при близких к нулю  $\alpha$ , выборочное значение статистики, меньшее чем  $x_\alpha$  выглядит неправдоподобно при нулевой гипотезе. Напротив, при  $\alpha$  близких к единице, значение, большее  $x_\alpha$  выглядит правдоподобно. Для построения таблицы критических точек будем для всех интересующих значений пользоваться следующей функцией с различными значениями параметров:

$$\text{calculate\_critical}(\text{rho\_statistic}, 3000000, T, 40, \alpha)$$

Здесь  $T$  – длина временного ряда,  $\alpha$  – ранее определённая вероятность. Числа  $3 \times 10^6$  и 40 подобраны так, чтобы с вероятностью 0,99 обеспечить погрешность вычислений, не превышающую  $10^{-3}$ .  $3 \times 10^6$  – это число реализаций, сгенерированных для расчёта. Основная погрешность для функции распределения определяется теоремой Колмогорова и равна  $\frac{1,63}{\sqrt{3 \times 10^6}} \cong 0,000941 < 0,001$ . Число 40 – это число итераций при приближённом решении уравнения  $P(T(\hat{\rho} - 1) < x_\alpha) = \alpha$  методом деления отрезка пополам. Погрешность такого приближения имеет тот же порядок, что и  $\frac{1}{2^{40}}$ . Такая погрешность на несколько порядков меньше основной, поэтому ей можно пренебречь без ущерба для точности. Таким образом, получаем значения критических точек, которые точны до двух знаков после запятой с вероятностью 0,99. Поскольку скорость сходимости эмпирической функции к теоретической невелика, дальнейшее увеличение точности сопряжено с большой вычислительной сложностью. Найдём, пользуясь функцией, критические значения для различных  $\alpha$  и нескольких длин ряда  $T$ . Отметим, что все вычисления здесь и далее производится при начальном условии  $y_0 = 0$  и единичной дисперсии белого шума. Однако, как уже было указано в теоретической части работы, распределение статистики не зависит от этих параметров, и их значения, вообще говоря, можно выбирать произвольно.

Результаты вычислений приводятся в таблице 1 ниже.

Критические точки распределения $p$ -статистики (таблица 1)										
$T$	$\alpha$									
	0,01	0,025	0,05	0,075	0,1	0,9	0,925	0,95	0,975	0,99
25	−11,88	−9,33	−7,35	−6,18	−5,34	1,02	1,19	1,42	1,80	2,31
50	−12,75	−9,87	−7,69	−6,42	−5,52	0,97	1,13	1,35	1,70	2,16
100	−13,22	−10,15	−7,86	−6,54	−5,62	0,95	1,11	1,32	1,65	2,09
150	−13,37	−10,26	−7,92	−6,59	−5,65	0,94	1,10	1,30	1,64	2,07
200	−13,48	−10,30	−7,94	−6,60	−5,67	0,94	1,09	1,30	1,63	2,06
250	−13,49	−10,32	−7,96	−6,61	−5,67	0,93	1,09	1,30	1,63	2,05

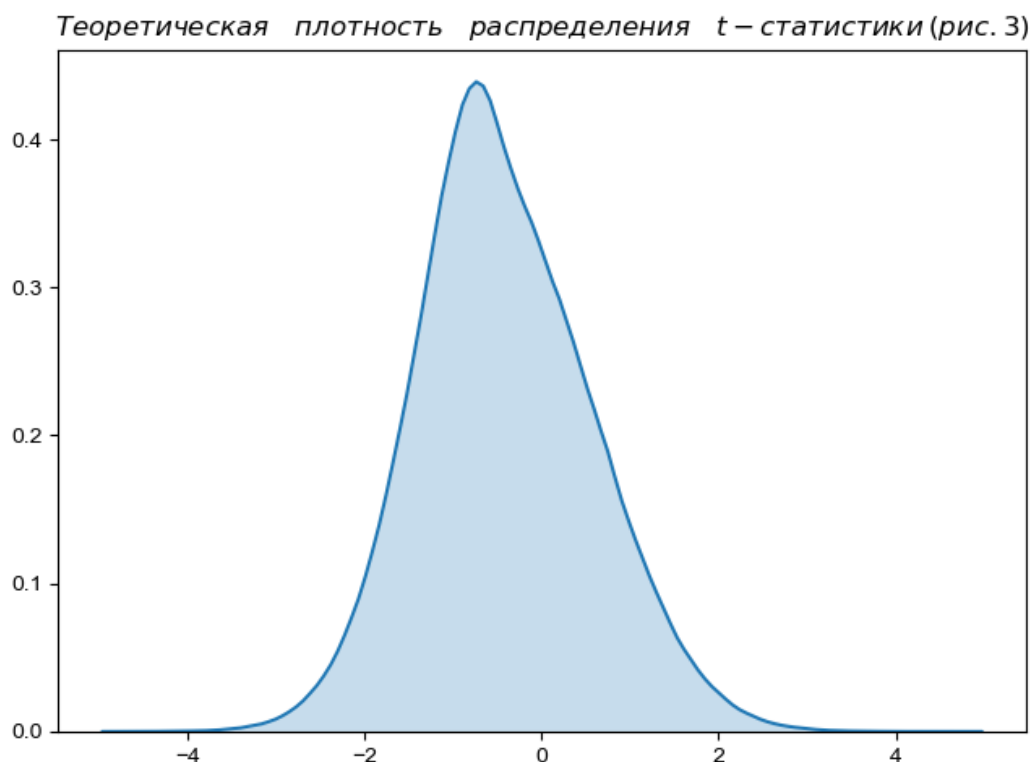
По столбцам таблицы можно проследить сходимость функции распределения к теоретической: значения критических точек убывают сверху вниз, притом есть тенденция к стабилизации. Отметим, что при  $T = 250$  число реализаций, генерируемых для расчёта уменьшено с  $3 \times 10^6$  до  $2,7 \times 10^6$  в связи с нехваткой оперативной памяти. Это не приводит к существенной потере точности:  $\frac{1,63}{\sqrt{2,7 \times 10^6}} \cong 0,000992 < 0,001$ . Для вычисления критических точек распределений при  $T > 300$  используемым методом с той же точностью требуется более 4 ГБ ОЗУ.

Составление таблиц критических точек достаточно трудоёмко. Не всегда для тестов на стационарность можно найти таблицу с подходящими  $T$  и  $\alpha$ , чтобы исследовать конкретный временной ряд. В таких случаях приходится руководствоваться значениями для ближайшего  $T$  из представленных в таблице. Однако, существует и альтернативный составлению таблиц подход, который часто более удобен – расчёт  $p$ -значения. Это значение для данной выборочной величины  $z$  тестовой статистики определяется как вероятность  $p = P(T(\hat{p} - 1) < z)$ . Фактически, происходит определение уровня значимости по заданным  $T$  и  $z$ . При этом понятно, что значение  $p$ , близкое к нулю говорит о неправдоподобности  $z$  при нулевой гипотезе. В приложении к работе есть функция для расчёта  $p$ -значения. Пример её применения будет рассмотрен в разделе 3.4.

### 3.3 Расчёты для $t$ -статистики

В этом разделе в полной аналогии с предыдущим производятся расчёты для  $t$ -статистики. Все выводы и оценки погрешности для  $\rho$ -статистики справедливы и здесь.

Построим график теоретической плотности  $DF_t$ . Необходимый массив реализаций генерируется функцией `calculate_theoretical(t_theoretical(1000000, 5000))`.



Эта плотность, конечно, отличается от плотности распределения  $\rho$ -статистики. Однако, поскольку она тоже построена на основании выражения  $\hat{\rho} - 1$  при нулевой гипотезе, наблюдается характерная скошенность распределения в сторону отрицательных значений.

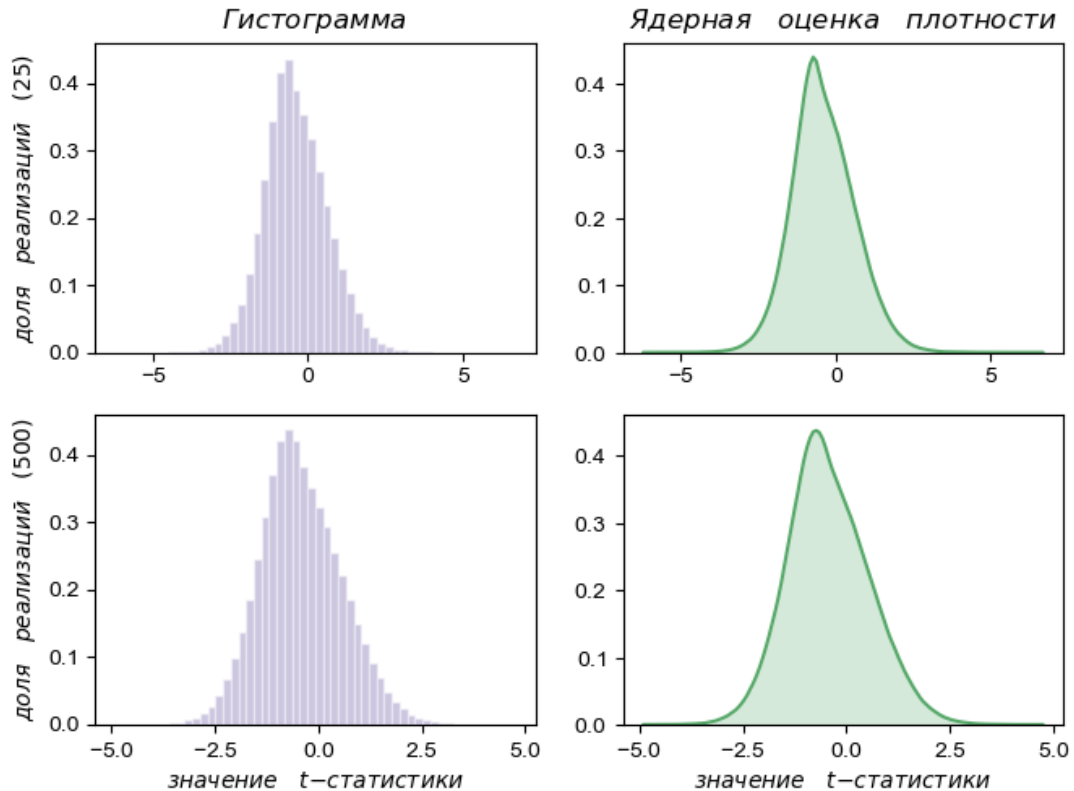
Графики плотностей для конечных объёмов выборки получим, как в предыдущем разделе на основе массивов реализаций меньшего объёма, чем нужно для вычисления критических точек:

`calculate_statistic(t_statistic(1000000, 25, 1, 0))`

`calculate_statistic(t_statistic(1000000, 500, 1, 0))`

Два последних параметра отвечают за дисперсию и начальное условие, соответственно.

Плотности распределения для рядов длины 25 и 500 (рис. 4)



На рисунке 4 видно, как плотность сближается с теоретической при  $T = 500$ .

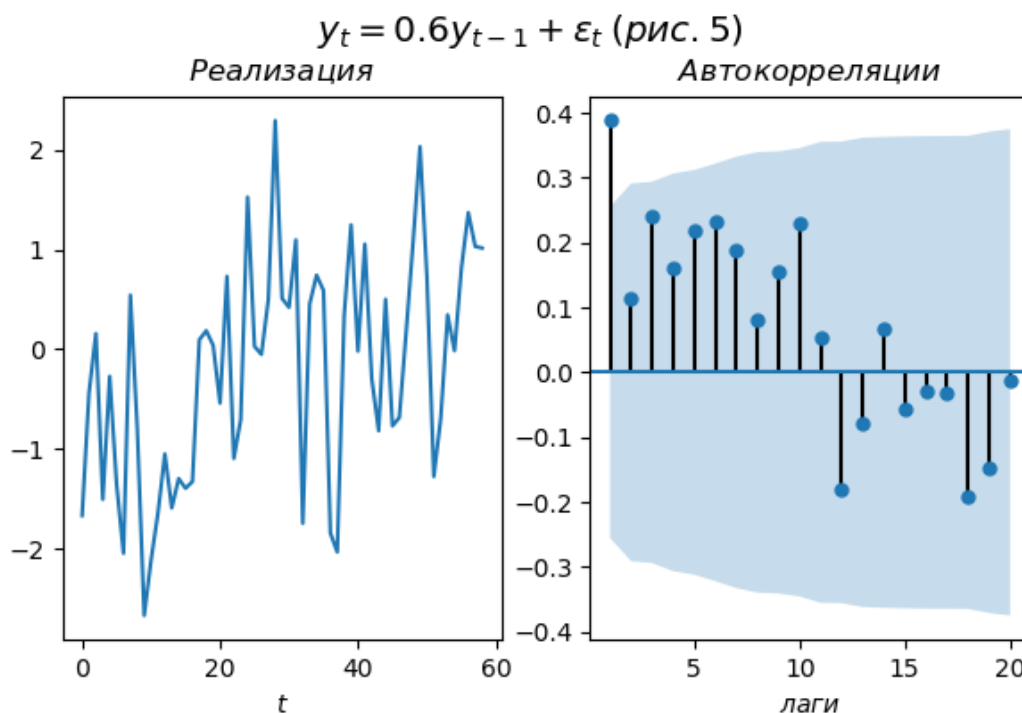
Приведём также таблицу критических точек, полученной при тех же ограничениях и с той же точностью, что и для  $r$ -статистики:

Критические точки распределения $t$ -статистики (таблица 2)										
$T$	$\alpha$									
	0,01	0,025	0,05	0,075	0,1	0,9	0,925	0,95	0,975	0,99
25	-2,66	-2,27	-1,95	-1,75	-1,61	0,93	1,11	1,34	1,71	2,14
50	-2,61	-2,25	-1,95	-1,76	-1,61	0,91	1,08	1,31	1,66	2,08
100	-2,59	-2,24	-1,94	-1,76	-1,62	0,90	1,07	1,30	1,64	2,04
150	-2,58	-2,24	-1,94	-1,76	-1,62	0,89	1,07	1,29	1,63	2,03
200	-2,58	-2,23	-1,94	-1,76	-1,62	0,89	1,07	1,29	1,63	2,03
250	-2,57	-2,23	-1,94	-1,76	-1,62	0,89	1,06	1,29	1,63	2,02

Соответствующая функция: `calculate_critical(t_statistic, 3000000, T, 40,  $\alpha$ )`.

### 3.4 Пример теста на стационарность

Приведём простой пример использования рассмотренных тестов. Для этого сгенерируем реализацию процесса авторегрессии первого порядка длины 59:



Поскольку временной ряд является реализацией искусственно сгенерированной стационарной авторегрессии первого порядка, тест должен дать основания для отклонения нулевой гипотезы. Рассчитаем статистики:

$$\rho_{statictic}(series) \rightarrow -34,1008555898639$$

$$t_{statictic}(series) \rightarrow -4,940866241777124$$

Сверясь со строками таблиц 1 и 2, соответствующими  $T = 50$ , ожидаемо получаем, что нулевую гипотезу можно отвергнуть с уровнем значимости 0,01.

Альтернативно, можем рассчитать  $p$ -значения, соответствующие выборочным статистикам, вызывая функцию  $p\_value(z, statistic, n, T)$ , где  $z$  – выборочное значение нужной статистики, а  $statistic$  – сама нужная функция-статистика. Погрешность по теореме Колмогорова с вероятностью 0,99 не превышает величины  $\frac{1,63}{\sqrt{n}}$ . Получаемые для данного примера  $p$ -значения крайне малы, что также говорит о неправдоподобности выборочных значений статистики при нулевой гипотезе.

## 4 Заключение

В данной работе рассмотрен простой тест Дики – Фуллера. Изложены его теоретические основания, приведены доказательства некоторых важных утверждений. Также проведён ряд расчётов: получены графики плотностей распределения и таблицы критических точек. Для этих расчётов с помощью теоремы Колмогорова произведена оценка погрешности. Приведён простой пример использования тестов. Работа соответствует принципу воспроизводимости, поскольку в приложении приведена программа, намеренно составленная так, чтобы быть наиболее универсальной и понятной.



## 5 Литература

1. Hayashi F. Econometrics – Princeton, 2000. – 669 p.
2. Теория вероятностей и математическая статистика: учеб. пособие / Л. Н. Фадеева, А. В. Лебедев; под ред. д-ра экон. наук, проф. Л. Н. Фадеевой – М.: Рид Групп, 2011. – 496 с. – (Национальное экономическое образование).
3. Теория вероятностей и математическая статистика: учебник и практикум для среднего профессионального образования / В. А. Малугин. – Москва: Издательство Юрайт, 2019. – 470 с. – (Профессиональное образование).
4. Вербик Марно Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Научн. ред. и предисл. С. А. Айвазяна – М.: Научная книга, 2008. – 616 с.
5. Носко В. П. Эконометрика. Кн. 1.4. 1,2: учебник / В. П. Носко. – М.: Издательский дом «Дело» РАНХиГС, 2011. – 672 с.

## 6 Приложение: программа на Python 3.7

```
# Импорт модулей

import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import math

# Функции

def white_noise(sigma, T, seed):
    """Реализация независимого белого шума длины T,  $\varepsilon \sim N(0, \sigma^2)$ """
    np.random.seed(seed)
    noise = np.random.normal(loc=0, scale=sigma, size=(1, T))
    return noise

def random_walk(initial_condition, noise):
    """Реализация случайного блуждания длины (T + 1) при начальном условии"""
    Y = initial_condition
    walk = np.append([Y], (Y + np.cumsum(noise)))
    return walk

def rho_statistic(series):
    """Вычисление  $\rho$ -статистики по временному ряду"""
    T = len(series) - 1
    series_diff = np.array(pd.Series(series).diff().dropna())
    series_ = np.array(series)[: -1]
    sum_upper = np.sum(series_*series_diff)
    sum_lower = np.sum(series_**2)
    rho_value = T*sum_upper/sum_lower
    print(rho_value)
    return rho_value

def t_statistic(series):
    """Вычисление t-статистики по временному ряду"""
    T = len(series) - 1
    series_diff = np.array(pd.Series(series).diff().dropna())
    series_ = np.array(series)[: -1]
    _series = np.array(series)[1:]
    term_upper = np.sum(series_*series_diff)
    rho = (rho_statistic(series)/T) + 1
    s = math.sqrt((1/(T-1))*np.sum((_series - rho*series_)**2))
    term_lower = s*math.sqrt(np.sum(series_**2))
    t_value = term_upper/term_lower
    print(t_value)
    return t_value

def calculate_statistic(statistic, N, M, sigma, Y):
    """Генерирование N случайных блужданий длины M с дисперсией белого шума
    sigma и начальным условием Y и подсчёт по ним статистики statistic"""
    noise_list = [white_noise(sigma, M-1, seed) for seed in range(N)]
    walk_list = [random_walk(Y, noise) for noise in noise_list]
    statistic_array = np.array([statistic(series) for series in walk_list])
    return statistic_array
```

```

def rho_theoretical(n, seed):
    """Приближённое вычисление значения одной теоретической реализации
    ρ-статистики с разбиением отрезка интегрирования на n равных частей"""
    sample = random_walk(0, white_noise(math.sqrt(1/n), n, seed))
    middles = (sample[:-1]+sample[1:])/2
    term_upper = (sample[-1]**2 - 1)/2
    term_lower = (np.sum(middles**2))/n
    rho_value = term_upper/term_lower
    print(rho_value)
    return rho_value

def t_theoretical(n, seed):
    """Приближённое вычисление значения одной теоретической реализации
    t-статистики с разбиением отрезка интегрирования на n равных частей"""
    np.random.seed(seed)
    sample = random_walk(0, white_noise(math.sqrt(1/n), n, seed))
    middles = (sample[:-1]+sample[1:])/2
    term_upper = (sample[-1]**2 - 1)/2
    term_lower = math.sqrt((np.sum(middles**2))/n)
    t_value = term_upper/term_lower
    print(t_value)
    return t_value

def calculate_theoretical(statistic, N, n):
    """Генерирование N реализаций теоретического распределения
    статистики statistic, где в каждой реализации отрезок
    интегрирования разбивается на n равных частей"""
    statistic_array = np.array([statistic(n, seed) for seed in range(N)])
    return statistic_array

def calculate_critical(statistic, N, M, I, alpha):
    """Генерирование N реализаций распределения статистики statistic,
    и вычисление критического значения для вероятности alpha"""
    array = calculate_statistic(statistic, N, M, 1, 0)
    left = np.min(array)
    right = np.max(array)
    x = (left + right)/2
    for _ in range(I):
        x_less = [k for k in array if k < x]
        P = len(x_less)/len(array)
        if P < alpha:
            left = x
            x = (left + right)/2
        elif P > alpha:
            right = x
            x = (left + right)/2
        else:
            break
    print(x)
    return x

def p_value(z, statistic, N, M):
    """Вычисление p-значения для величины z статистики statistic"""
    array = calculate_statistic(statistic, N, M, 1, 0)
    x_less = [k for k in array if k < z]
    p = len(x_less)/len(array)
    print(p)
    return p

```