



Национальный исследовательский университет «Высшая школа
экономики»

Факультет: Московский институт электроники и математики

Образовательная программа: Прикладная математика

Отчет по Самостоятельной работе № 1

Временные ряды и их практическое применение

Работу выполнили:
Цыплаков Александр
Ермаков Семен
4 апреля 2024 г.

Содержание

Содержание.....	2
Введение.....	3
Исходные данные.....	4
Основные компоненты временного ряда.....	6
3.1.....	6
3.2.....	7
3.3.....	8
Методология Бокса-Дженкинса.....	11
4.1.....	11
4.2.....	13
4.3.....	14
Выбранные модели для Франции.....	14
Выбранные модели для США.....	16
4.4.....	18
4.5.....	22
4.6.....	24
Заключение.....	25
Приложения.....	26

Введение

В современном мире энергетические ресурсы играют ключевую роль в обеспечении экономического роста и социального развития. Одним из важных аспектов энергетической политики является производство электроэнергии из ядерных источников. В этом контексте проведен анализ и прогнозирование временных рядов производства электроэнергии из ядерных источников в Соединенных Штатах Америки и Франции.

Выбор данных Соединенных Штатов и Франции обусловлен их значимостью на мировой арене и различиями в стратегиях развития ядерной энергетики. Период наблюдений охватывает временной интервал с 1960 по 2015 год, что позволяет проанализировать долгосрочную динамику развития ядерной энергетики в этих странах.

Исходные данные

Анализируемый показатель и его единица измерения:	Производство электроэнергии из ядерных источников (% от общего объема)
Анализируемые страны/регионы:	Соединенные Штаты Америки и Франция
Источник данных:	Данные Всемирного банка (https://data.worldbank.org/indicator?tab=all)
Период наблюдения:	1960 - 2015 годы
Количество наблюдений:	N = 56 (за каждый год с 1960 по 2015)



Рис. 1. Распределение данных

Согласно графикам, видно, что в период с 1960-х по 1990 годы Франция пережила резкий скачок в производстве электроэнергии из ядерных источников. Этот скачок указывает на успешную демонстрацию эффективности ядерной энергетики в стране и переход на использование этого источника энергии. В то же время, Соединенные Штаты Америки характеризовались более равномерным ростом производства электроэнергии из ядерных источников на протяжении всего исследуемого периода.

Тем не менее, несмотря на различия в динамике роста между двумя странами (США: 27793.18%; Франция: 39808.67%), важно отметить, что обе они демонстрируют значительные показатели роста в производстве электроэнергии из ядерных источников. Темп прироста в обеих странах также подтверждает эту динамику (США: 0.1058; Франция: 0.1129), хотя темп

прироста для Франции немного выше, чем для Соединенных Штатов Америки. В целом, общая динамика производства электроэнергии из ядерных источников за рассматриваемый период показывает важность и эффективность этого источника энергии для обеих стран.

Дескриптивные статистики:

Полученные дескриптивные статистики для производства электроэнергии из ядерных источников в Соединенных Штатах Америки и Франции представляют следующую картину:

Соединенные Штаты Америки:

Среднее значение: 13.06%
Стандартное отклонение: 7.70%
Минимальное значение: 0.07%
Максимальное значение: 20.65%

Франция:

Среднее значение: 47.71%
Стандартное отклонение: 33.76%
Минимальное значение: 0.19%
Максимальное значение: 79.51%

Из этой информации видно, что в среднем Франция имеет значительно более высокий процент производства электроэнергии из ядерных источников, чем Соединенные Штаты Америки. Стандартное отклонение для Франции также значительно выше, что указывает на более высокую изменчивость в производстве электроэнергии из ядерных источников в этой стране по сравнению с США. В то время как минимальное и максимальное значения указывают на разнообразие ситуаций в обеих странах в отношении использования ядерной энергии.

Основные компоненты временного ряда

3.1

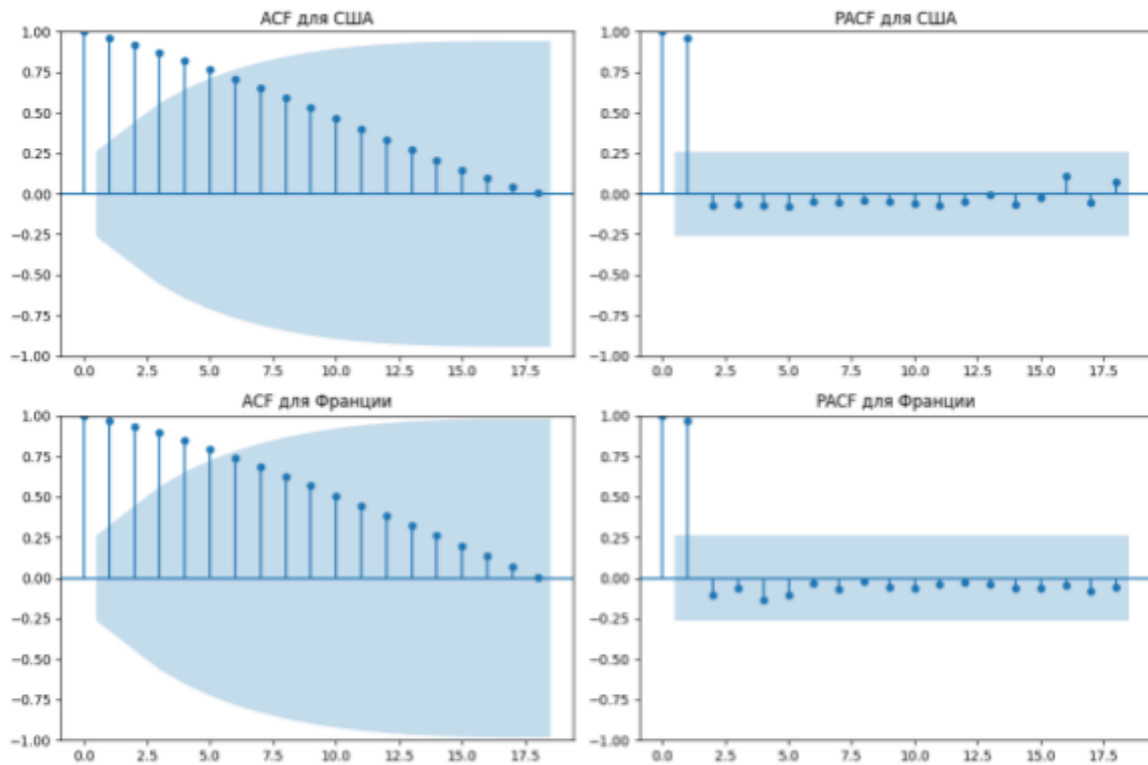


Рис. 2. Графики ACF и PACF для США и Франции

На изображении представлены четыре графика, отображающие автокорреляционную функцию (ACF) и частичную автокорреляционную функцию (PACF) для двух разных датасетов, а именно для США (верхние графики) и Франции.

1. Верхний левый график ("ACF для США") показывает автокорреляционную функцию для США. Автокорреляция измеряется для разных лагов (задержек), и мы видим, что значения автокорреляции положительны и медленно уменьшаются по мере увеличения лага. Это указывает на то, что в данных присутствует некоторая зависимость от предыдущих значений, и она уменьшается по мере увеличения разрыва во времени.

2. Верхний правый график ("PACF для США") показывает частичную автокорреляционную функцию для США. Значения PACF в основном

находятся внутри границ синего облака, что означает, что после учёта предыдущих лагов, большинство оставшихся корреляций не значительны. Однако, первый лаг выходит за границы этого облака, что указывает на статистически значимую частичную автокорреляцию для этого лага.

3. Нижний левый график ("ACF для Франции") аналогично показывает автокорреляционную функцию, но для Франции. Здесь также видна положительная автокорреляция, которая уменьшается с увеличением лага, что также указывает на временную зависимость в данных.

4. Нижний правый график ("PACF для Франции") показывает частичную автокорреляционную функцию для Франции. Здесь также видим, что большинство значений PACF находится внутри границ, за исключением первого лага, который существенно выделяется, указывая на его значимость.

Графики содержат синее облако вокруг оси лагов, отображающее доверительный интервал (95%), внутри которого значения корреляции не являются статистически значимыми. Это предполагает, что значения выходящие за пределы этого облака считаются статистически значимыми.

Данные временные ряды скорее всего не являются стационарными, поскольку имеется значительная корреляция как в ACF, так и в PACF. При этом наличие высоких значений на первом лаге в PACF может свидетельствовать о наличии структурных изменений в данных или сезонности. Также из анализа ACF и PACF можно сделать вывод, что временные ряды для США и Франции обладают трендовой компонентой, что подтверждается положительной автокорреляцией и наличием значимых лагов на графиках PACF. Тренд характеризуется увеличением производства электроэнергии из ядерных источников со временем.

3.2

Для проверки гипотезы об отсутствии тренда во временных рядах можно использовать тест Квадратичного Приращения Параметров (KPSS).

Результаты теста представлены следующим образом:

Для временного ряда США:

- KPSS Statistic: 1.059654
- p-value: 0.010000

Для временного ряда Франции:

- KPSS Statistic: 1.072875
- p-value: 0.010000

Исходя из результатов теста KPSS, нулевая гипотеза о стационарности временного ряда отвергается при уровне значимости 0.01, так как p-value меньше уровня значимости. Это означает, что у нас есть статистически значимые доказательства в пользу наличия тренда в обоих временных рядах.

Выводы:

На основании результатов теста Квадратичного Приращения Параметров (KPSS) обнаружено, что как для временного ряда Соединенных Штатов Америки, так и для Франции, нулевая гипотеза о стационарности временного ряда была отвергнута на уровне значимости 0.01. Это свидетельствует о наличии тренда в обоих временных рядах.

Эти результаты подтверждают выводы, сделанные ранее при анализе графиков временных рядов. Мы отметили, что в обоих странах присутствует тренд, проявляющийся в изменении уровня производства электроэнергии из ядерных источников с течением времени. Таким образом, наличие тренда в данных подтверждается как визуальным анализом графиков, так и результатами статистического теста.

3.3

На графике CUSUM для США (USA CUSUM) мы наблюдаем, что кривая начинается ниже нулевой линии и постепенно поднимается к ней. Это означает, что в начале рассматриваемого периода данные имеют тенденцию к убыванию или меньшему значению, чем ожидалось.

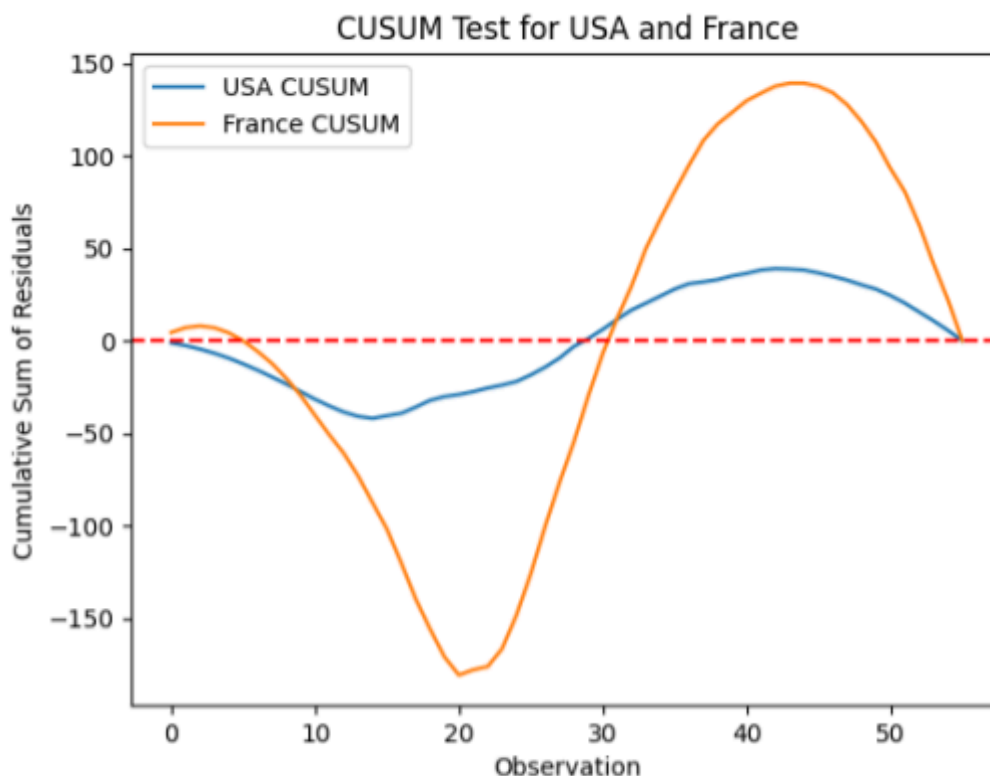


Рис. 3. CUSUM-тест

Затем кривая пересекает нулевую линию и начинает восходящее движение, достигая максимальной амплитуды в районе значения 40. Это указывает на сдвиг в данных, который может быть связан с каким-то значимым событием или изменением в экономике или политике США.

Аналогично, график CUSUM для Франции (France CUSUM) также начинается ниже нулевой линии и имеет более выраженную амплитуду, достигая около значения 150. Это указывает на более сильный структурный сдвиг в данных Франции по сравнению с США.

Эти выводы подтверждаются результатами теста Чоу (Chow test). Для Соединенных Штатов Америки значение F-статистики составляет примерно 20.52, а p-value составляет приблизительно 2.69×10^{-7} . Это означает, что полученное значение F-статистики статистически значимо, а p-value крайне мало, что позволяет нам отвергнуть нулевую гипотезу об отсутствии структурного сдвига в данных для США.

Для Франции значение F-статистики составляет около 15.45, а p-value около 5.41×10^{-6} . Эти значения также свидетельствуют о статистически значимом структурном сдвиге в данных для Франции.

Таким образом, результаты теста Чоу подтверждают наличие структурных изменений во временных рядах производства электроэнергии из ядерных источников как для Соединенных Штатов Америки, так и для Франции.

Теперь мы можем составить следующую таблицу:

	США	Франция
Наличие тренда	Да, статистика критерия: 1.059654, p-value: 0.010000, статистически значимый тренд в данных	Да, статистика критерия: 1.072875, p-value: 0.010000, статистически значимый тренд в данных
Наличие структурного сдвига	Да, F-статистика: 20.52, p-value: 2.69e-07, статистически значимый сдвиг в данных	Да, F-статистика: 15.45, p-value: 5.41e-06, статистически значимый сдвиг в данных

Возможные причины структурных сдвигов в данных о производстве электроэнергии из ядерных источников в США и Франции.

- **Технологические инновации:** Возможно, в рассматриваемый период в одной из стран произошел значительный технологический прорыв в области ядерной энергетики. Это могло привести к увеличению производства электроэнергии из ядерных источников и, как следствие, к структурному сдвигу в данных.
- **Энергетическая политика:** Изменения в энергетической политике могли повлиять на развитие ядерной энергетики. Например, введение новых стимулов или регулирований в отрасли может существенно повлиять на объем производства электроэнергии из ядерных источников.
- **Экологические факторы:** Рост общего интереса к экологии и усиление требований к сокращению выбросов парниковых газов могли стать стимулом для развития чистых источников энергии, включая ядерную энергетику.
- **Экономические факторы:** Экономические изменения, такие как изменение стоимости топлива или экономические кризисы, могли повлиять на инвестиции в ядерную энергетику и соответствующее производство электроэнергии.
- **Геополитические события:** События мирового масштаба, такие как изменения в международных отношениях, конфликты или сокращение поставок энергоносителей, могли повлиять на стратегии энергетической безопасности и внутреннюю политику по использованию ядерной энергии.

Учитывая тему исследования "производство электроэнергии из ядерных источников", структурный сдвиг в данных может быть связан с одним или

комбинацией вышеперечисленных факторов, которые оказывают влияние на развитие ядерной энергетики в рассматриваемых странах.

Методология Бокса-Дженкинса

4.1

Для проверки стационарности временного ряда мы использовали тесты ADF (Augmented Dickey-Fuller), PP (Phillips-Perron) и KPSS (Kwiatkowski-Phillips-Schmidt-Shin). Они проверяют нулевую гипотезу о наличии единичного корня в ряде, что указывает на нестационарность. В случае тестов ADF и PP, если p -значение меньше выбранного уровня значимости (обычно 0.05), нулевая гипотеза отвергается, и ряд считается стационарным. Напротив, в тесте KPSS, если статистика теста меньше критического значения, нулевая гипотеза о стационарности отвергается, что указывает на нестационарность ряда.

Достоинства и недостатки тестов:

Тест Дики-Фуллера: Преимущества включают его широкое использование и простоту интерпретации. Однако он менее мощный и может давать ложноположительные результаты.

Тест Филлипса-Перрона: Этот тест более устойчив к некоторым типам автокорреляции и может быть более мощным. Недостатком является зависимость результатов от выбора параметров.

Тест KPSS: Он хорошо работает для обнаружения нестационарности ряда. Однако он менее чувствителен к некоторым типам структурных изменений и может давать ложноположительные результаты

При анализе временных рядов с использованием статистических тестов, о которых было сказано ранее, мы столкнулись с противоречивыми, что создает неопределенность в определении стационарности временного ряда. Противоречия заключались в том, что после применения метода последовательных разностей (мпр) некоторые тесты для Франции, а именно: ADF (с трендом и без) и KPSS (с трендом) утверждали о нестационарности данных, хотя тесты PP (с трендом и без) и KPSS (без тренда) утверждали о стационарности, вследствие чего было принято решение применить еще раз мпр, чтобы все тесты согласовывались. Такое решение было принято ввиду последующих заданий, которые требовали стационарные данные и нам нужно было быть уверенными в этом. После применения мпр ко Франции во второй раз, все тесты показали стационарность.

Порядок интегрируемости процесса определяет, сколько разностей необходимо применить к временному ряду, чтобы сделать его стационарным.

Далее приведена сводная таблица о тестах и их результатах для обеих стран

США			
ВР	Тест	Статистика критерия, р-значение	Вывод
у	ADF (без тренда)	-1.7739, 0.3935	Нестационарный
	ADF (с трендом)	-0.2005, 0.9915	Нестационарный
	PP (без тренда)	0.647, 0.856	Нестационарный
	PP (с трендом)	-0.473, 0.984	Нестационарный
	KPSS (без тренда)	1.0597, 0.01	Нестационарный
	KPSS (с трендом)	0.2763, 0.01	Нестационарный
Δу	ADF (без тренда)	-5.8017, 4.6156e-07	Стационарен
	ADF (с трендом)	-6.111, 1.2798e-06	Стационарен
	PP (без тренда)	-6.013, 0.000	Стационарен
	PP (с трендом)	-6.047, 0.000	Стационарен

	KPSS (без тренда)	0.4167, 0.0699	Стационарен
	KPSS (с трендом)	0.123, 0.0925	Стационарен
Франция			
BP	Тест	Статистика критерия, р-значение	Вывод
у	ADF (без тренда)	-1.6884, 0.4371	Нестационарный
	ADF (с трендом)	-2.2325, 0.4716	Нестационарный
	PP (без тренда)	0.491, 0.823	Нестационарный
	PP (с трендом)	-1.153, 0.920	Нестационарный
	KPSS (без тренда)	1.0729, 0.01	Нестационарный
	KPSS (с трендом)	0.2094, 0.0125	Нестационарный
Δу	ADF (без тренда)	-1.8259, 0.3677	Нестационарный
	ADF (с трендом)	-1.9854, 0.6094	Нестационарный
	PP (без тренда)	-5.220, 0.000	Стационарен
	PP (с трендом)	-5.465, 0.000	Стационарен
	KPSS (без тренда)	0.261, 0.1	Стационарен
	KPSS (с трендом)	0.1699, 0.0301	Нестационарный
Δ(Δу)	ADF (без тренда)	-10.4292, 1.6275e-18	Стационарен
	ADF (с трендом)	-10.3479, 4.3382e-16	Стационарен
	PP (без тренда)	-18.306, 0.000	Стационарен
	PP (с трендом)	-18.282, 0.000	Стационарен
	KPSS (без тренда)	0.1559, 0.1	Стационарен
	KPSS (с трендом)	0.1229, 0.0926	Стационарен

Табл. 1. Результаты тестов на стационарность для США и Франции

Общий вывод: Исходя из результатов проведенного анализа, можно сделать вывод о том, что тип процесса временного ряда для США может быть определен как стационарный (Stationary TS) с порядком интегрируемости равным 1, в то время как для Франции тип процесса такой же, но порядок интегрируемости равен 2.

4.2

ARIMA (Autoregressive Integrated Moving Average) - это распространенный метод моделирования временных рядов, который включает в себя авторегрессионную (AR), интегрированную (I) и скользящее среднее (MA) компоненты. Выбор конкретной ARIMA-модели зависит от анализа автокорреляционной функции (ACF) и частичной автокорреляционной функции (PACF) временного ряда.

Существует несколько типов ARIMA-моделей:

- $ARIMA(p, d, q)$: Это базовая модель ARIMA, где p - порядок авторегрессии, d - порядок интегрирования, q - порядок скользящего среднего.
- $ARIMA(p, d, 0)$: Это модель $AR(p)$ или авторегрессионная модель, которая учитывает только предыдущие значения временного ряда.
- $ARIMA(0, d, q)$: Это модель $MA(q)$ или модель скользящего среднего, которая учитывает только предыдущие ошибки прогнозирования.
- $ARIMA(p, 0, q)$: Это модель $ARMA(p, q)$, которая не включает интегрирование и использует как авторегрессионные, так и скользящие средние компоненты.

Для подбора параметров ARIMA-модели на практике часто используют методы перебора, например, с помощью сетки параметров (grid search) или метода наименьших квадратов (least squares). Также широко используются информационные критерии, такие как AIC (Akaike Information Criterion) и BIC (Bayesian Information Criterion), для сравнения различных моделей ARIMA и выбора наилучшей из них с учетом баланса между точностью и сложностью модели.

4.3

Выбранные модели для Франции

1. ARIMA(3, 2, 3)

Математическая форма модели:

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3)y_t = \varepsilon_t$$

Статистические свойства оценок:

- $\phi_1 = -0.3310, \phi_2 = -0.2543, \phi_3 = -0.0871$ (коэфф. авторегрессии)
- $\theta_1 = -0.4171, \theta_2 = 0.1354, \theta_3 = 0.4155$ (коэфф. скользящего среднего)
- $\sigma^2 = 7.5878$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 268.460
- BIC: 282.119
- Ljung-Box (Q): 0.02 (не отвергается гипотеза о неавтокоррелированности остатков)

Предпосылки ARIMA моделей:

- Стационарность: В модели нет явных признаков нестационарности, но необходимо дополнительно провести тесты на стационарность остатков.
- Обратимость: Обратимость модели подтверждается.

2. ARIMA(2, 2, 1)

$$(1 - \phi_1 L - \phi_2 L^2)(1 - \theta_1 L)y_t = \varepsilon_t$$

Статистические свойства оценок:

- $\phi_1 = -0.7042, \phi_2 = -0.4917$ (коэфф. авторегрессии)
- $\theta_1 = -0.0015$ (коэфф. скользящего среднего)
- $\sigma^2 = 8.1756$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 265.638
- BIC: 273.443
- Ljung-Box (Q): 0.01 (не отвергается гипотеза о неавтокоррелированности остатков)

Предпосылки ARIMA моделей:

- Стационарность: Требуется дополнительная проверка на стационарность остатков
- Обратимость: Обратимость модели подтверждается.

3. ARIMA(1, 2, 0)

Математическая форма модели: $(1 - \phi_1 L)(1 - L^2)y_t = \varepsilon_t$

Статистические свойства оценок:

- $\phi_1 = -0.4755$ (коэфф. авторегрессии)
- $\sigma^2 = 10.9228$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 276.150
- BIC: 280.053
- Ljung-Box (Q): 3.26 (не отвергается гипотеза о неавтокоррелированности остатков)

Предпосылки ARIMA моделей:

- Стационарность: Требуется дополнительная проверка на стационарность остатков
- Обратимость: Обратимость модели подтверждается.

4. Модель с детерминированным трендом ARIMA(1, 2, 0):

Математическая форма модели:

$$(1 - \phi_1 L)(1 - L^2)y_t = c + \varepsilon_t$$

Параметры:

- $c = 4.292 \times 10^{-8}$
- $\phi_1 = -0.4755$ (коэфф. авторегрессии)
- $\sigma^2 = 10.9228$

Критерии качества модели:

- AIC: 278.150
- BIC: 284.004
- Ljung-Box (Q): 3.26 (не отвергается гипотеза о неавтокоррелированности остатков)

5. Модель с фиктивной переменной ARIMA(1, 2, 0):

Математическая форма модели:

$$(1 - \phi_1 L)(1 - L^2)y_t = \beta_1 D_{1t} + \varepsilon_t$$

Параметры:

- $\beta_1 = 0$ (в наших данных нет структурного сдвига)
- $\phi_1 = -0.4755$ (коэфф. авторегрессии)
- $\sigma^2 = 10.9228$

Критерии качества модели:

- AIC: 278.150

- BIC: 284.004
- Ljung-Box (Q): 3.26 (не отвергается гипотеза о неавтокоррелированности остатков)

Выбранные модели для США

1. ARIMA(1, 1, 1)

Математическая форма модели: $(1 - \phi_1 L)(1 - \theta_1 L)y_t = \varepsilon_t$

Статистические свойства оценок:

- $\phi_1 = 0.9080$ (коэфф. авторегрессии)
- $\theta_1 = -0.7336$ (коэфф. скользящего среднего)
- $\sigma^2 = 7.5539$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 141.755
- BIC: 147.666
- Ljung-Box (Q): 0.01 (не отвергается гипотеза о неавтокоррелированности остатков)

Предпосылки ARIMA моделей:

- Стационарность: В модели нет явных признаков нестационарности, но необходимо дополнительно провести тесты на стационарность остатков.
- Обратимость: Обратимость модели подтверждается.

2. ARIMA(1, 1, 0)

Математическая форма модели: $(1 - \phi_1 L)y_t = \varepsilon_t$

Статистические свойства оценок:

- $\phi_1 = 0.3181$ (оценки параметров авторегрессии)
- $\sigma^2 = 0.8156$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 143.711
- BIC: 147.651
- Ljung-Box (Q): 0.97 (не отвергается гипотеза о неавтокоррелированности остатков)

Предпосылки ARIMA моделей:

- Стационарность: Требуется дополнительная проверка на стационарность остатков
- Обратимость: Обратимость модели подтверждается.

3. ARIMA(3, 1, 3)

Математическая форма модели:

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3)y_t = \varepsilon_t$$

Статистические свойства оценок:

- $\phi_1 = 1.2221, \phi_2 = -1.1681, \phi_3 = 0.6065$ (коэфф. авторегрессии)
- $\theta_1 = -1.0752, \theta_2 = 1.2586, \theta_3 = -0.3746$ (коэфф. скользящего среднего)
- $\sigma^2 = 0.6463$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 144.886
- BIC: 158.678
- Ljung-Box (Q): 0.03 (не отвергается гипотеза о неавтокоррелированности остатков)

Предпосылки ARIMA моделей:

- Стационарность: Требуется дополнительная проверка на стационарность остатков
- Обратимость: Обратимость модели подтверждается.

4. Модель с детерминированным трендом ARIMA(1, 1, 0):

Математическая форма модели:

$$(1 - \phi_1 L)(1 - L^2)y_t = c + \varepsilon_t$$

Параметры:

- $c = 0.3569$
- $\phi_1 = 0.2074$ (коэфф. авторегрессии)
- $\sigma^2 = 0.7471$ (оценка дисперсии ошибок)

Критерии качества модели:

- AIC: 141.002
- BIC: 146.913
- Ljung-Box (Q): 0.01 (не отвергается гипотеза о неавтокоррелированности остатков)

5. Модель с фиктивной переменной ARIMA(1, 1, 0):

Математическая форма модели:

$$(1 - \phi_1 L)(1 - L^2)y_t = \beta_1 D_1 + \varepsilon_t$$

Параметры:

- $\beta_1 = 0$ (в наших данных нет структурного сдвига)
- $\phi_1 = 0.3181$

- $\sigma^2 = 0.8156$

Критерии качества модели:

- AIC: 145.711
- BIC: 151.621
- Ljung-Box (Q): 0.97 (не отвергается гипотеза о неавтокоррелированности остатков)

4.4

США				
Модели ARIMA	Стационарность, обратимость	Ошибка модели	Анализ остатков	Наилучшая модель
ARIMA(1, 1, 1)	стационарна, обратима	AIC=141.755 BIC=147.666	p-value: 0.089 Skew: -0.478 Kurtosis: 4.113	
ARIMA(1, 1, 0)	стационарна, обратима	AIC=143.711 BIC=147.651	p-value: 0.575 Skew: 0.014 Kurtosis: 3.701	
ARIMA(3, 1, 3)	стационарна, обратима	AIC=144.886 BIC=158.678	p-value: 0.0004 Skew: -0.779 Kurtosis: 5.088	
ARIMA(1, 1, 0) deterministic_trend	стационарна, обратима	AIC=141.002 BIC=146.913	p-value: 0.538 Skew: 0.1302 Kurtosis: 3.696	✓
ARIMA(1, 1, 0) dummy_shock	стационарна, обратима	AIC=145.711 BIC=151.621	p-value: 0.575 Skew: 0.014 Kurtosis: 3.701	
Франция				
Модели ARIMA	Стационарность, обратимость	Ошибка модели	Анализ остатков	Наилучшая модель
ARIMA(3, 2, 3)	стационарна, обратима	AIC=268.460 BIC=282.119	p-value: 0.005 Skew: -0.339 Kurtosis: 5.055	
ARIMA(2, 2, 1)	стационарна, обратима	AIC=265.638 BIC=273.443	p-value: 0.025 Skew: 0.133	✓

			Kurtosis: 4.789	
ARIMA(1, 2, 0)	стационарна, обратима	AIC=276.150 BIC=280.053	p-value: 0.176 Skew: -0.029 Kurtosis: 4.241	
ARIMA(1, 2, 0) deterministic_trend	стационарна, обратима	AIC=278.150 BIC=284.004	p-value: 0.176 Skew: -0.029 Kurtosis: 4.241	
ARIMA(1, 2, 0) dummy_shock	стационарна, обратима	AIC=145.711 BIC=151.621	p-value: 0.176 Skew: -0.029 Kurtosis: 4.241	

Все тесты Дики-Фуллера показывают, что значения p-value близки к нулю для всех моделей как для Франции, так и для США. Это говорит о том, что мы можем отвергнуть нулевую гипотезу о наличии единичных корней (не стационарность) в остатках моделей. Таким образом, все модели имеют стационарные остатки.

За нормальность распределения остатков отвечает значение p-value, если оно меньше 5%, то мы отвергаем нулевую гипотезу о распределении остатков, если же результат превышает пороговое значение, то мы не можем отвергнуть нулевую гипотезу и следовательно остатки распределены нормально

Skewness (скос) и kurtosis (куртозис) - это две меры, используемые для характеристики формы распределения данных.

Skewness (скос):

- Скос показывает, насколько данные смещены относительно нормального распределения.
- Если скос равен нулю, это означает, что данные симметричны относительно среднего значения.
- Если скос положителен, это указывает на то, что хвост распределения смещен вправо относительно среднего значения, что означает наличие более тяжелого правого хвоста.
- Если скос отрицателен, это указывает на то, что хвост распределения смещен влево относительно среднего значения, что означает наличие более тяжелого левого хвоста.

Kurtosis (куртозис):

- Куртозис показывает, насколько данные остры или плоски.
- Если куртозис равен 3, это означает, что данные имеют "нормальный" пик.

- Значение куртозиса больше 3 указывает на более острые пики (более концентрированные данные вокруг среднего), а значения меньше 3 указывают на более плоские пики (более разбросанные данные относительно среднего).

Таким образом, Skew и Kurtosis помогают оценить форму распределения данных и обнаружить отклонения от нормальности.

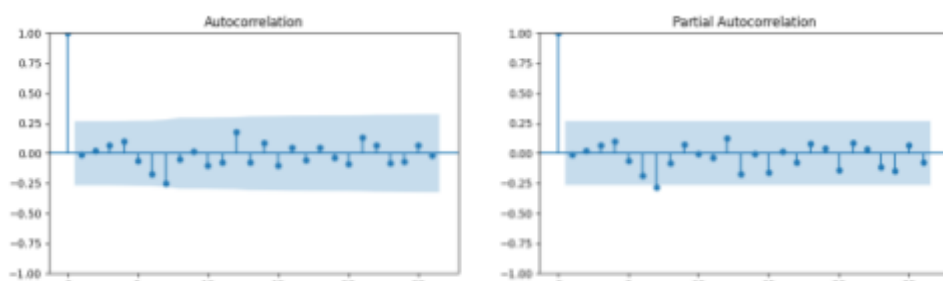


Рис. 4. Результаты автокорреляции для Франции. Модель ARIMA(2,2,1).

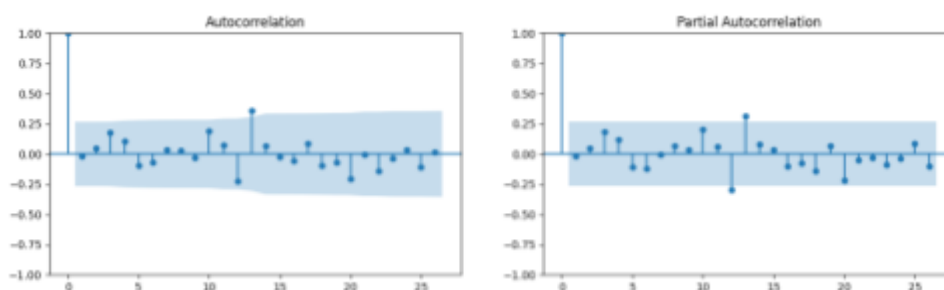


Рис. 5. Результаты автокорреляции для США.
Модель ARIMA(1,1,0) deterministic_trend

На графиках автокорреляции для данных Франции видны индивидуальные точки, каждая из которых представляет коэффициент автокорреляции для соответствующего лага (задержки).

Большинство коэффициентов автокорреляции находятся внутри синего облака, которое обозначает доверительный интервал, и лишь одна точка для первого лага несколько выше верхней границы, что может указывать на наличие статистически значимой автокорреляции для этого лага. Однако все последующие показатели автокорреляции находятся внутри границ и близки к нулю, что говорит об отсутствии значимой автокорреляции на этих уровнях лагов.

На графике частичной автокорреляции также представлены точки, показывающие величину автокорреляции для каждого лага с учетом влияния всех более коротких лагов. Похоже, что все точки также находятся внутри доверительного интервала, что свидетельствует об отсутствии значимой частичной автокорреляции для всех представленных лагов.

Вывод: Оба графика указывают на то, что в остатках временного ряда отсутствует значимая автокорреляция, за исключением возможной автокорреляции на первом лаге, видной на графике автокорреляции. Это

предполагает, что временной ряд может быть моделирован без необходимости учета автокорреляции. Для США ситуация аналогична.

Исходя из оценок AIC, BIC лучшей моделью для каждой страны будет:

- Для США:
Лучшая модель: ARIMA(1, 1, 0) с детерминированным трендом
 - AIC: 141.002
 - BIC: 146.913
- Для Франции:
Лучшая модель: ARIMA(2, 2, 1)
 - AIC: 265.638
 - BIC: 273.443

Эти модели являются лучшими, при этом значения p-value, Skew, Kurtosis свидетельствуют о нормальности распределения остатков, что соответствует заданным нами критериям модели.

4.5

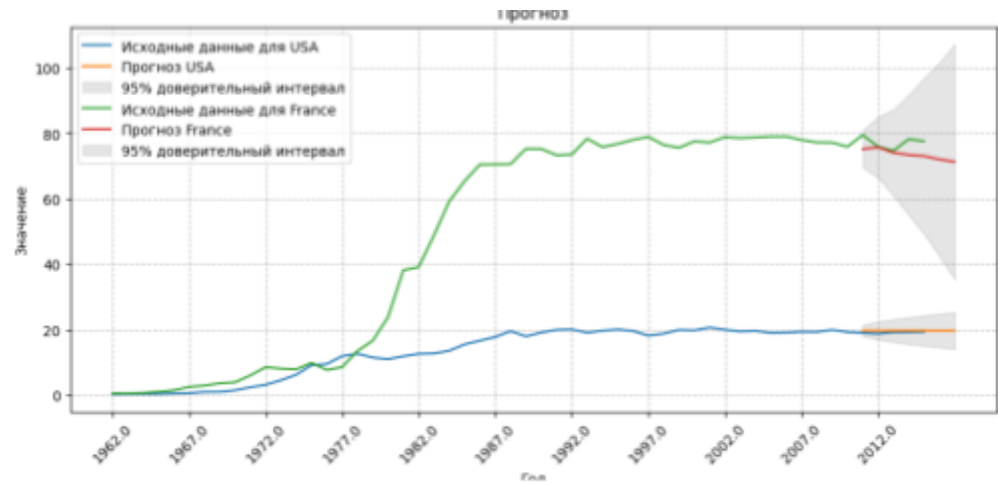


Рис.6. Прогноз значений для обеих стран

Исходя из полученных значений предсказаний и фактических данных для страны США, можно отметить следующее:

Для США фактические значения электроэнергии и предсказанные точечные значения в течение следующих шести периодов представлены следующим образом:

Год	Фактические значения	Предсказанные точечные значения
-----	----------------------	---------------------------------

2011	18.984846	19.730024
2012	18.757920	19.704282
2013	19.173831	19.698914
2014	19.230239	19.697795
2015	19.323778	19.697562
2016	-	19.697513
2017	-	19.697503

Из представленных данных видно, что предсказанные значения электроэнергии в целом близки к фактическим значениям, что может свидетельствовать о хорошей точности модели.

Аналогично для Франции:

Год	Фактические значения	Предсказанные точечные значения
2011	79.511660	75.227544
2012	75.962780	75.770294
2013	74.704486	74.138064
2014	78.235588	73.443350
2015	77.627801	73.084509
2016	-	72.063816
2017	-	71.352569

Также был построен 95% доверительный интервал для каждой страны, информация о которых представлена в следующих таблицах:

Для США

Год	Нижняя граница	Верхняя граница
2011	17.989623	21.470424
2012	16.974301	22.434263
2013	16.205978	23.191851
2014	15.572578	23.823013
2015	15.023305	24.371819
2016	14.531959	24.863068
2017	14.083427	25.311580

Для Франции:

Год	Нижняя граница	Верхняя граница
2011	69.514803	80.940285
2012	66.442888	85.097700
2013	60.992957	87.283170
2014	55.069115	91.817585
2015	49.281856	96.887162
2016	42.489444	101.638187
2017	35.386104	107.319034

Полученные доверительные интервалы позволяют оценить уверенность прогноза для каждой страны. Верхняя и нижняя границы представляют собой диапазон значений, в котором с определенной вероятностью (95%) ожидается нахождение фактических значений переменной в будущем.

Наблюдаемая ширина интервала может свидетельствовать о степени неопределенности прогноза: чем шире интервал, тем меньше уверенность в прогнозе. Для наших данных ширина доверительного интервала достаточно большая, что свидетельствует о плохой уверенности в прогнозе.

4.6

Оценка качества прогноза для США:

- MSE: 0.32253013714127576
- MAE: 0.5225157083470374
- RMSE: 0.5679173682335096
- R-squared: -1.8675630752888677

Оценка качества прогноза для Франции:

- MSE: 17.04058713470872
- MAE: 3.44496965550653
- RMSE: 4.128024604421432
- R-squared: -6.580092985187145

Результаты оценки качества модели показывают, что прогноз для США имеет более низкие показатели ошибок, чем для Франции. MSE и RMSE для США значительно меньше, чем для Франции, что указывает на более точные прогнозы для американской экономики. Однако оба прогноза характеризуются отрицательным значением коэффициента детерминации (R-squared), что может свидетельствовать о недостаточной объяснительной способности модели или наличии систематических ошибок в прогнозах.

Заключение

Результаты исследования показывают, что доля объема производства электроэнергии из ядерных источников от общего объема производства выходит на плато в последние годы в обеих странах. Прогнозирование с помощью ARIMA моделей позволяет понять, что ожидать роста этого процента в ближайшие годы не стоит (если не скажутся какие-то новые внешние факторы, например экономические или технические). При этом, видно, что доверительные интервалы весьма широкие, особенно для Франции, хотя бурного роста в будущем вряд ли стоит ожидать (согласно правилу 20/80). Следует подчеркнуть, что одним из потенциальных методов сокращения размера доверительного интервала является расширение

пространства признаков, а также применение прогностических моделей с многомерным входом.

Приложения

<https://github.com/ErmakovSemen/minor>