Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____________«ФУНДАМЕНТАЛЬНЫЕ НАУКИ__________________

КАФЕДРА __________________« ВЫСШАЯ МАТЕМАТИКА»____________________

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К КУРСОВОЙ РАБОТЕ

НА ТЕМУ:

Решение задач прикладной математики численными методами

Студент ____ФН1-71Б____ (Группа) __________________ (Подпись, дата) _**В.Д. Богословский**__ (И.О.Фамилия)

Руководитель курсовой работы __________________ (Подпись, дата) ____**А.А. Федотов**____ (И.О.Фамилия)

Консультант __________________ (Подпись, дата) ____**Л.Г. Ветров**_____ (И.О.Фамилия)

2020 г.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ
Заведующий кафедрой ___ФН1____
(Индекс)
______________ __Н.И. Сидняев__
(И.О.Фамилия)
« _11_ » _сентября___ 2020 г.

З А Д А Н И Е
на выполнение курсовой работы

по дисциплине _____________практикум по применению математических пакетов______________

Студент группы ___ ФН1-71Б ____

__________________________Богословский Владислав Дмитриевич________________________
(Фамилия, имя, отчество)

Тема курсовой работы __ Решение задач прикладной математики численными методами __

Направленность КР (учебная, исследовательская, практическая, производственная, др.)
__учебная, исследовательская ____
Источник тематики (кафедра, предприятие, НИР) __кафедра ФН1__________________________

График выполнения работы: 25% к _3_нед., 50% к _9_ нед., 75% к _12_ нед., 100% к _15_ нед.

Задание __ анализ и моделирование временных рядов в среде MATLAB

Оформление курсовой работы:

Расчетно-пояснительная записка на _____ листах формата А4.

Дата выдачи задания « _4_ » __сентября___ 2020г.

Руководитель курсовой работы _________________ ___ **А.А. Федотов** ___
(Подпись, дата) (И.О.Фамилия)

Студент _________________ _ **В.Д. Богословский**
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Оглавление

Введение

Временной ряд — собранный в разные моменты времени статистический материал о значении каких-либо параметров исследуемого процесса. Каждая единица статистического материала называется измерением или отсчётом. Временной ряд существенно отличается от простой выборки данных, так как при анализе учитывается взаимосвязь измерений со временем, а не только статистические характеристики выборки.

Анализ временных рядов — совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования. Выявление структуры временного ряда необходимо для того, чтобы построить математическую модель того явления, которое является источником анализируемого временного ряда. Прогноз будущих значений временного ряда используется для эффективного принятия решений.

Цели анализа временных рядов:

- определение природы ряда;
- прогнозирование (предсказание будущих значений временного ряда по настоящим и прошлым значениям).

Временные ряды, как правило, возникают в результате измерения некоторого показателя. Это могут быть как показатели (характеристики) технических систем, так и показатели природных, социальных, экономических и других систем (например, погодные данные). Типичным примером временного ряда можно назвать биржевой курс, при анализе которого пытаются определить основное направление развития (тенденцию или тренд).

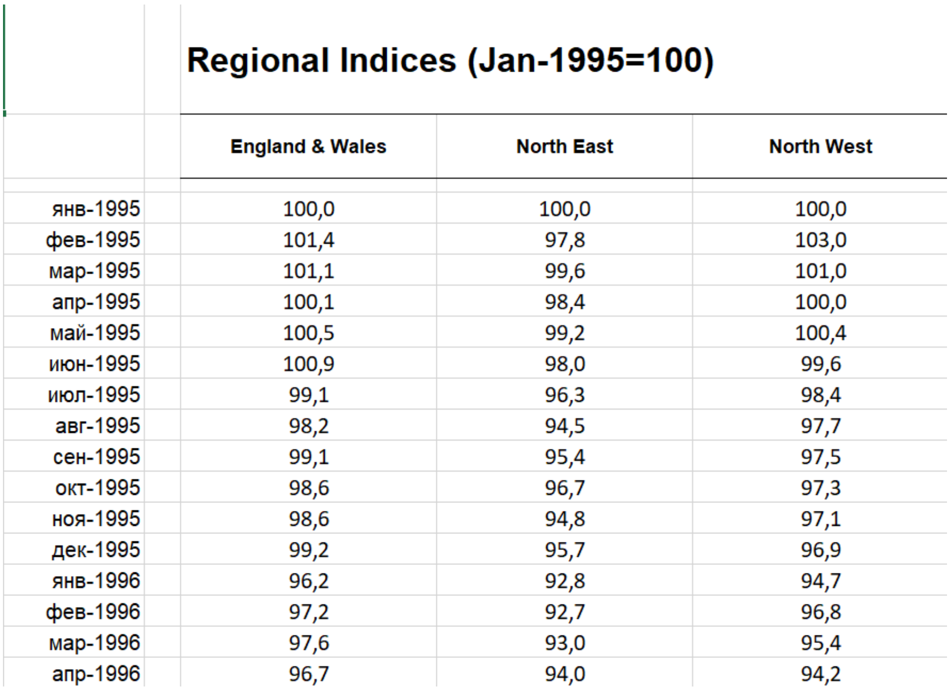
В данной работе мы рассмотрим моделирование временного ряда на примере индексов цен на жилье, а также произведем анализ коррелированности двух рядов. При написании курсовой работы использован математический пакет MATLAB, который имеет многочисленный набор инструментов (в том числе эконометрических), необходимых для решения данной задачи.

Глава 1

Задача 1: Моделирование временного ряда по моделе Бокса-Дженкинса

1.1 Постановка задачи

Дан временной ряд идексов цен на жилье за период с 1995г. по 2014г. в формате HPI.xls.



Regional Indices (Jan-1995=100)

	England & Wales	North East	North West
янв-1995	100,0	100,0	100,0
фев-1995	101,4	97,8	103,0
мар-1995	101,1	99,6	101,0
апр-1995	100,1	98,4	100,0
май-1995	100,5	99,2	100,4
июн-1995	100,9	98,0	99,6
июл-1995	99,1	96,3	98,4
авг-1995	98,2	94,5	97,7
сен-1995	99,1	95,4	97,5
окт-1995	98,6	96,7	97,3
ноя-1995	98,6	94,8	97,1
дек-1995	99,2	95,7	96,9
янв-1996	96,2	92,8	94,7
фев-1996	97,2	92,7	96,8
мар-1996	97,6	93,0	95,4
апр-1996	96,7	94,0	94,2

Рис. 1.1: Фрагмент исходных данных

Импортируя временной период и числовые показатели Англии и Уэльса, смоделировать поведение временного ряда, спрогнозировать его поведение на последующие 10 месяцев и провести тестирование методом Монте Карло.

1.2 Описание методологии прогнозирования

Временные ряды состоят из двух элементов:

- периода времени, за который или по состоянию на который приводятся числовые значения;
- числовых значений того или иного показателя, называемых уровнями ряда.

Классификация временных рядов:

- по количеству показателей (одномерные и многомерные временные ряды);
- по наличию пропущенных значений (полные и неполные временные ряды);
- в зависимости от наличия основной тенденции (стационарные и нестационарные ряды).

Наиболее распространенной причиной нарушения стационарности является тенденция в среднем значении, которое может быть обусловлено либо наличием единого корня, либо детерминированного тренда. В первом случае (DS-ряды) случайные скачки имеют постоянные эффекты, и процесс не является возвратным к среднему. В последнем случае (TS-ряды) скачки имеют только временные эффекты, после которых переменная стремится к детерминистически развивающемуся (непостоянному) среднему значению.

Стационарный относительно тренда ряд (*TS-ряд, trend stationary*) — временной ряд, для которого существует некоторая детерминированная функция, такая что их разность является стационарным процессом.

$$X_t \text{ – TS-series } \overset{def}{\Longleftrightarrow} \exists f(t) : X_t - f(t) \text{ - стационарный ряд.}$$

Важным видом нестационарного ряда, который не включает трендоподобное поведение, является **циклостационарный ряд**, который является стохастическим процессом, который циклически изменяется со временем.

Интегрированный временной ряд — нестационарный временной ряд, разности некоторого порядка от которого являются стационарным временным рядом. Такие ряды также называют *разностно-стационарными (DS-рядами, Difference Stationary).*

$$X_t \sim I(k) \overset{def}{\iff} \Delta^k X_t \sim I(0), \quad I(0) \text{ - стационарный ряд.}$$

Понятие интегрированного временного ряда тесно связано с единичными корнями в авторегрессионных моделях. Наличие единичных корней в характеристическом полиноме авторегрессионной составляющей модели временного ряда означает интегрированность временного ряда. Причем количество единичных корней совпадает с порядком интегрированности.

ARIMA (*autoregressive integrated moving average*) — модель авторегрессионного интегрированного скользящего среднего, являющаяся расширением моделей ARMA для нестационарных временных рядов. Модель $ARIMA(p,d,q)$ означает, что разности временного ряда порядка d подчиняются модели $ARMA(p,q)$.
Формальное определение $ARIMA(p,d,q)$ модели:

$$a(L)(1-L)^d X_t = c + b(L)\varepsilon_t.$$

Формальное определение $ARMA(p,q)$ модели:

$$a(L)X_t = c + b(L)\varepsilon_t,$$

<u>Замечание 1:</u> в обоих случаях $\Theta(L) = \sum\limits_{i=0}^{N} \theta_i L^i$ - лаговый многочлен.

<u>Замечание 2:</u> **лаговый оператор** (L)— оператор смещения, позволяющий получить значения элементов временного ряда на основании ряда предыдущих значений.

$$L^k X_t = X_{t-k}.$$

Этапы моделирования временного ряда в методологии Бокса-Дженкенса:

1. **Предобработка данных**

   В первую очередь оценивается стационарность ряда. Для интегрированных временных рядов применяется подход ARIMA. Различными тестами выявляются наличие единичных корней и порядок интегрированности временного ряда (обычно ограничиваются первым или вторым порядком). Далее при необходимости (если порядок интегрированности больше нуля) ряд преобразуется взятием разности соответствующего порядка и уже для преобразованной модели строится некоторая ARMA-модель.

   Тесты на стационарность:

   - Тест Квятковского–Филлипса–Шмидта–Шина (*KPSS test*);

     Проверяет гипотезу стационарности относительно тренда против альтернативной гипотезы о наличии единичного корня.

     Статистический критерий: $\frac{\sum\limits_{t=1}^{n} S_t^2}{s^2 n^2}$, где $S_t = e_1 + e_2 + \cdots + e_t$ (невязки МНК), n - размер выбоки, s^2 - оценка Ньюи-Уэста долгосрочной дисперсии (количество лагов автоковариации, которые необходимо включить в оценку Ньюи-Уэста предлагается считать равным$\sqrt{N}$).

   - Расширенный тест Дики–Фуллера (*ADF test*).

     Проверяет гипотезу наличия единичного корня против альтернативной о его отсутствии. Критические значения DF-статистики берутся из специальной таблицы Дики–Фуллера. Если значение статистики лежит левее критического значения (критические значения — отрицательные) при данном уровне значимости, то нулевая гипотеза о единичном корне отклоняется и процесс признается стационарным (в смысле данного теста). В противном случае гипотеза не отвергается и процесс может содержать единичные корни, то есть быть нестационарным (интегрированным) временным рядом.

   В случае TS-рядов для того, чтобы сделать ряд стационарным, нужно удалить входящие в него тренды, сезонные компоненты и бизнес циклы. О наличие сезонной составляющей можно судить по наличию периодически повторяющихся пиков автокорреляционной функции.

Для удаления этой состовляющей можно воспользоваться цифровой обработкой сигнала.

Для вычленения циклов можно использовать фильтр Ходрика–Прескотта. Он разделяет временной ряд на трендовые и циклические компоненты ($y_t = T_t + C_t$). Математически фильтр Ходрика-Прескотта записывается как задача минимизации следующей функции потерь:

$$\sum_{t=1}^{N}(y_t - T_t)^2 + \lambda \sum_{t=2}^{N-1}((T_{t+1} - T_t) - (T_t - T_{t-1}))^2 \to \min.$$

Рекомендуемые параметры сглаживания λ: 14400 (месяц), 1600 (квартал), 100 (год).

Первая сумма минимизирует разницу между временным рядом и его трендовой компонентой (которая является его циклической составляющей). Вторая сумма минимизирует разницу второго порядка тренда (что аналогично минимизации второй производной тренда).

2. **Выбор модели**

   На основании автокорреляционной (ACF) и частной автокорреляционной функции ($PACF$) определятся порядок членов авторегрессии и скользящего среднего.

   Автокорреляция — статистическая взаимосвязь между последовательностями величин одного ряда, взятыми со сдвигом.

   Автокорреляция k-го порядка:

$$r_k = \frac{\frac{1}{n}\sum\limits_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\hat{\sigma}^2(y_t)}.$$

   Автокорреляционная функция показывает зависимость автокорреляции от величины сдвига во времени $r_k = r(k)$. **Частная** же **автокорреляционная функция** показывает прямое влияние y_t на y_{t-k} (исключая влияние $y_{t-1}, \ldots, y_{t-k+1}$).

   Оптимальная модель выбирается на основе информационного критерия Акаике.

   Информационный критерий — применяемая в статистике мера относительного качества статистических моделей, учитывающая степень «подгонки» модели под данные с корректировкой на используемое количество оцениваемых параметров. То есть критерии основаны на неком компромиссе между точностью и сложностью модели. Критерии различаются тем, как они обеспечивают этот баланс. Чем меньше значение критерия, тем лучше модель.

Информационный критерий Акаике рассчитывается по формуле

$$AIC = 2k - 2l,$$

где l - значение логарифмической функции правдоподобия построенной модели, а k - количество использованных (оцененных) параметров.

3. **Оценка модели**

   На данном этапе производится оценка остаточного распределения, проверка на гетероскедастичность и окончательная оценка параметров комбинированной ARIMA/GARCH модели (*Generalized AutoRegressive Conditional Heteroscedasticity*).

   Тесты на нормальный закон распределения:

   - Тест Жака-Бера

     Двусторонний критерий согласия, проверяющий гипотезу о том, что данные распределены нормально при неизвестных среднем и дисперсии. Тест проверяет ошибки наблюдений на нормальность посредством сверки их третьего и четвертого момента с моментами нормального распределения, у которого асимметрия и эксцесс равны соответственно $s = 0$ и $k = 3$.

     Соответственно проверяется нулевая гипотеза $H_0 : s = 0, k = 3$ против альтернативной $H_1 : s \neq 0, k \neq 3$.

     Тестовая статистика:

     $$JB = \frac{n}{6}\left(\hat{s}^2 + \frac{(\hat{k} - 3)^2}{4}\right),$$

     где n – размер выборки, $\hat{s}$ – асимметрия выборки, а $\hat{k}$ – эксцесс выборки. Для больших размеров выборки статистика теста имеет распределение χ^2 с двумя степенями свободы.

- Критерий Андерсона — Дарлинга

  Критерий согласия Андерсона — Дарлинга предназначен для проверки простых гипотез о принадлежности анализируемой выборки полностью известному закону, то есть для проверки гипотез вида $H_0 : F_n(x) = F(x, \theta)$ с известным вектором параметров теоретического закона.

  Тестовая статистика принадлежит к семейству статистик с квадратичными эмпирическими функциями распределения, которые измеряют расстояние между гипотетической $F(x)$ и эмпирической $F_n(x)$ функциями как $n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 w(x) dF(x)$ по элементам вариационного ряда $x_{(1)} < \cdots < x_{(n)}$, где n-количество точке в данной выборке, а $w(x) = [F(x)(1 - F(x))]^{-1}$ – весовая функция, которая придает наблюдениям больший вес в хвостах распределения, тем самым делая тест более чувствительным к выбросам и лучше обнаруживая отклонения от нормальности в хвостах распределения.

  Тестовая статисика Андерсона-Дарлинга:

  $$A_n^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln F(x_i) + \ln(1 - F(x_{n+1-i})) \right],$$

  где $\{x(1) < \cdots < x_n\}$ – вариационный ряд, а n – количество точек данных в выборке.

- Критерий Лиллиефорса

  Критерий Лиллиефорса — статистический критерий, являющийся модификацией критерия Колмогорова–Смирнова. Используется для проверки нулевой гипотезы о том, что выборка распределена по нормальному закону для случая, когда параметры нормального распределения априори неизвестны.

  Тестовая статистика Лиллиефорса:

  $$D^* = \max |F_n(x) - G(x)|,$$

  где $F_n(x)$ – эмпирическая функция распределения выборки, а $G(x)$ – гипотетическая функция распределения с оцененными параметрами.

  Основным источником погрешности критерия Лиллиефорса является то обстоятельство, что параметры теоретического распределения оцениваются по тем же самым данным, которые проверяются на соответствие распределению. Таким образом, максимальное отклонение будет меньше, чем в случае, когда параметры распределения оцениваются независимо. Поэтому «нулевое распределение» статистики критерия, т.е. распределение вероятности в предположении об истинности нулевой гипотезы, оказывается смещено в сторону меньших значений по сравнению с распределением Колмогорова. Оно известно как «распределение Лиллиефорса» и рассчитывается методом Монте-Карло.

Критерий случайности остатков:

Q-тест Льюнга-Бокса – это тест, который оценивает нулевую гипотезу о том, что ряд остатков не демонстрирует автокорреляции для фиксированного числа лагов L, против альтернативы, что некоторый коэффициент автокорреляции $\rho(k)$, $k = 1, \ldots, L$, отличен от нуля.

Тестовая статистика:

$$Q = n(n+2)\sum_{k=1}^{L}\frac{\hat{\rho}^2(k)}{n-k},$$

где n – размер выборки, L – количество лагов автокорреляции, а $\hat{\rho}(k)$ – автокорреляция выборки порядка k.

При нулевой гипотезе асимптотическое распределение Q является распределением χ^2 с L степенями свободы. Если $Q > \chi^2_{1-\alpha}(L)$, где $\chi^2_{1-\alpha}(L)$ – квантили распределения χ^2 с L степенями свободы, то нулевая гипотеза отвергается, и признаётся наличие автокорреляции до L-го порядка во временном ряду.

Гетероскедастичность — неоднородность наблюдений, выражающаяся в неодинаковой (непостоянной) дисперсии случайной ошибки модели. Гетероскедастичность противоположна *гомоскедастичности*, означающей однородность наблюдений, то есть постоянство дисперсии случайных ошибок модели.

Тест на условную гетероскедастичность:

ARCH-тест Энгла оценивает нулевую гипотезу о том, что ряд остатков r_t не проявляет условной гетероскедастичности (ARCH-эффекты), в отличие от альтернативы, согласно которой модель $ARCH(L)$ описывает ряд.

$ARCH(L)$ модель имеет следующий вид:

$$r_t^2 = a_0 + \sum_{i=1}^{L} a_i r_{t-i}^2 + \varepsilon_t,$$

где есть хотя бы одно $a_j \neq 0$, $j = 0, \ldots, L$.

Тестовая статистика – это статистика множителей Лагранжа nR^2, где n – размер выборки, а R^2 – коэффициент детерминации при подборе модели $ARCH(L)$ для числа лагов L с помощью регрессии.

При нулевой гипотезе асимптотическое распределение тестовой статистики представляет собой χ^2 с L степенями свободы.

4. **Симуляция и прогноз**

   Создается прогноз с доверительными границами, рассчитываются траектории модели Монте-Карло для стационарных данных с извлечением средней траектории и ее доверительных интервалов и выполняется обратное преобразование (возвращение всех удаленных компонент и кумулятивное суммирование).

1.3 Решение поставленной задачи

1.3.1 Предобработка данных

Импортируем данные из таблицы HPI.xls

```
T = readtable('HPI.xls','Sheet','Indices', ...
    'Range','B4:D233', 'ReadVariableNames',false);
dates = T.Var1;
HPI = T.Var3;
```

Выведем график.

```
plot(dates, HPI)
title('Индексы цен на жилье, 1995-2014')
xlabel('Даты')
ylabel('HPI')
grid on
```

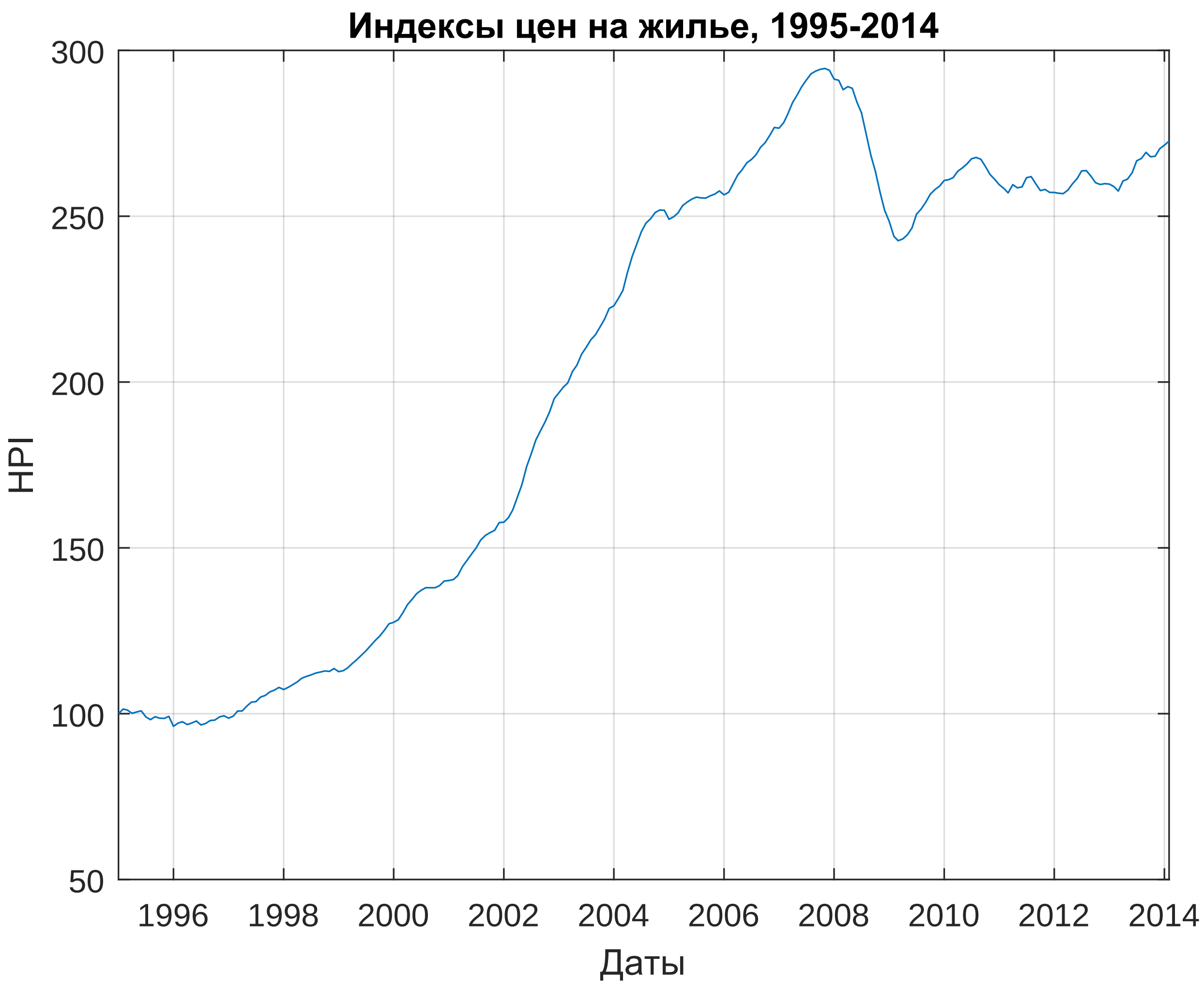


Рис. 1.2: Исходный временной ряд

Проанализируем основные характеристики статистических данных - среднее значение и среднеквадратичное отклонение (далее СКО).

```
n_fragments = rem(numel(HPI), 12); % Количество дней в последнем неполном месяце
years = year(dates(1:end-n_fragments)); % Года с повторениями
yrsHPI = HPI(1:end-n_fragments);
[meanHPIByYear, stdHPIByYear] =...
    grpstats(yrsHPI, years, {@mean, @std}); % Среднее значение и ско HPI по годам
yrs = unique(years); % Года
```

Визуализация соотношения среднего и СКО.

```
figure('Name', 'Характеристики статистических данных', 'NumberTitle', 'off');
subplot(2,1,1)
plot(yrs, meanHPIByYear, 'r.-', 'LineWidth', 1.5)
title('Среднегодовые значения HPI')
xlabel('Year')
ylabel('HPI')
grid on
subplot(2,1,2)
plot(yrs, stdHPIByYear, 'LineWidth', 1.5)
title('Среднегодовые значения СКО HPI')
xlabel('Year')
ylabel('HPI')
grid on
```

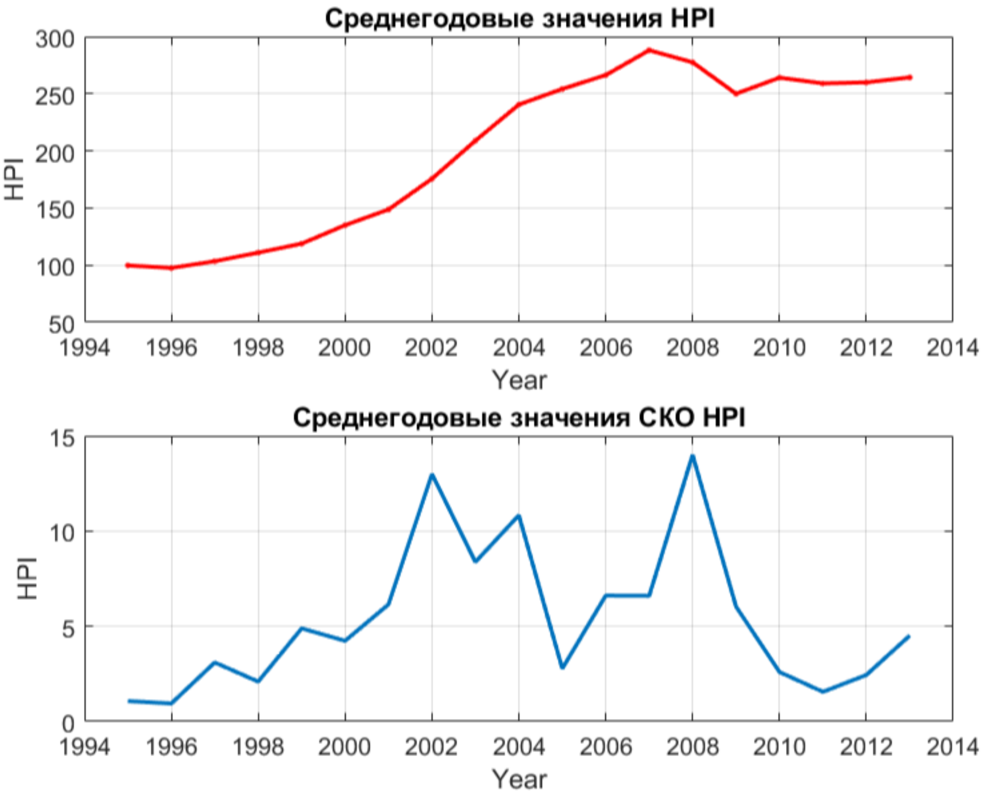


Рис. 1.3: Характеристики статистичеких данных

Приходим к выводу, что данные нестационарны.

Выявление экпоненциального тренда

Для данных, имеющих в своем составе паттерн с экспоненциальным поведением, справедлива линейная зависимость математического ожидания и среднего квадратичного отклонения, так как $\xi \sim Exp(\lambda) \Rightarrow$ $m_1 = \sigma = \frac{1}{\lambda}$.

```
figure
scatter(meanHPIByYear,stdHPIByYear,'kx')
xlabel('Среднее HPI')
ylabel('СКО HPI')
title('Соотношение среднего и СКО для HPI')
grid on
l = lsline;
set(l, 'Color', 'r', 'LineWidth', 2.5)
[c, p] = corr(meanHPIByYear, stdHPIByYear, 'type', 'Spearman', 'tail', 'right');
display_text = char('Коэффициент корреляции и p-value:', num2str([c, p]));
text( min(meanHPIByYear), mean(stdHPIByYear)+std(stdHPIByYear), display_text)
```

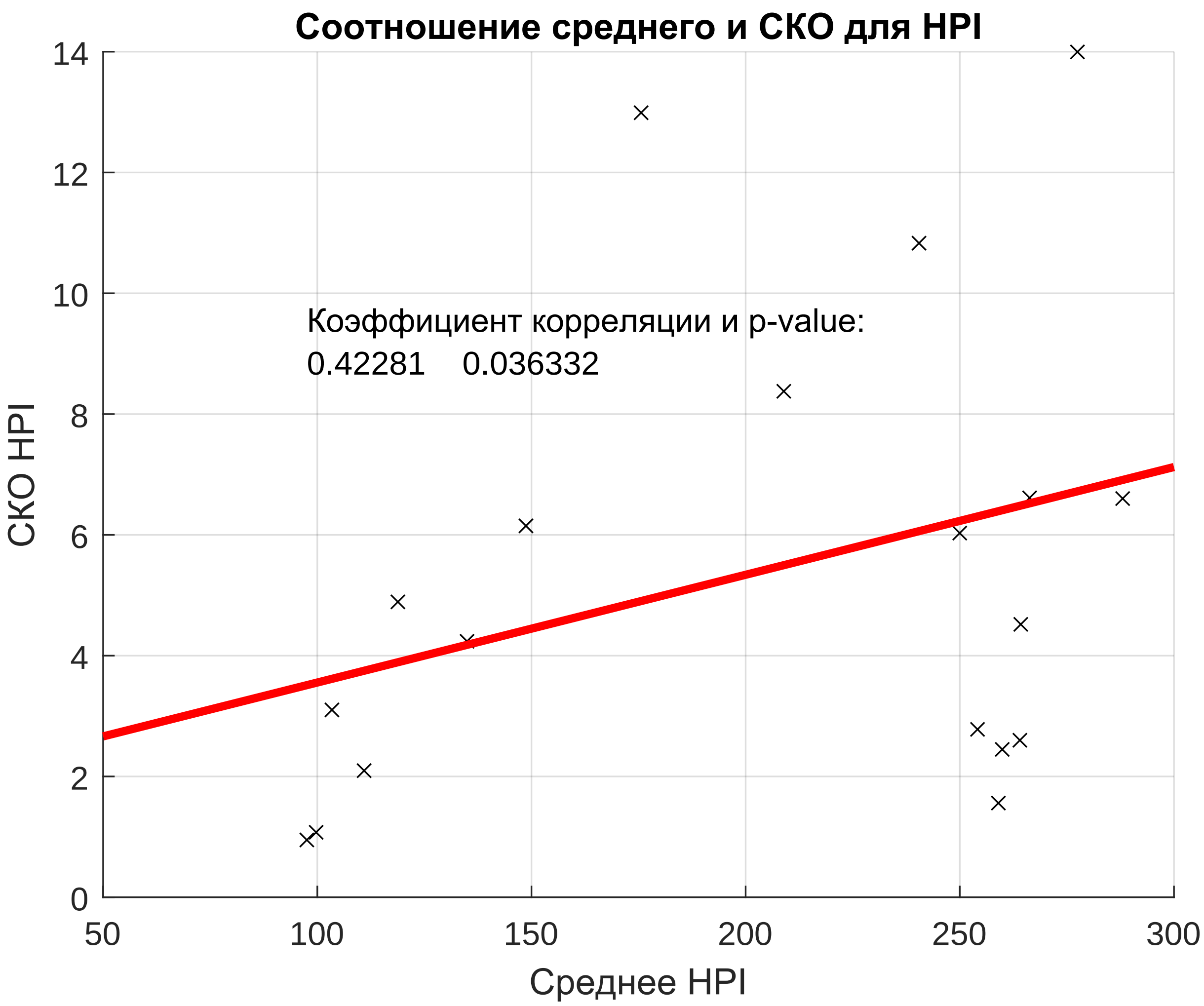


Рис. 1.4: Выявление экпоненциального тренда

Удаление экспоненциального тренда из HPI серии:

```
HPI_log = log(HPI);
plot(dates, HPI_log*100/HPI_log(1))
title('HPI серия с удаленным экспоненциальным трендом')
grid on
```

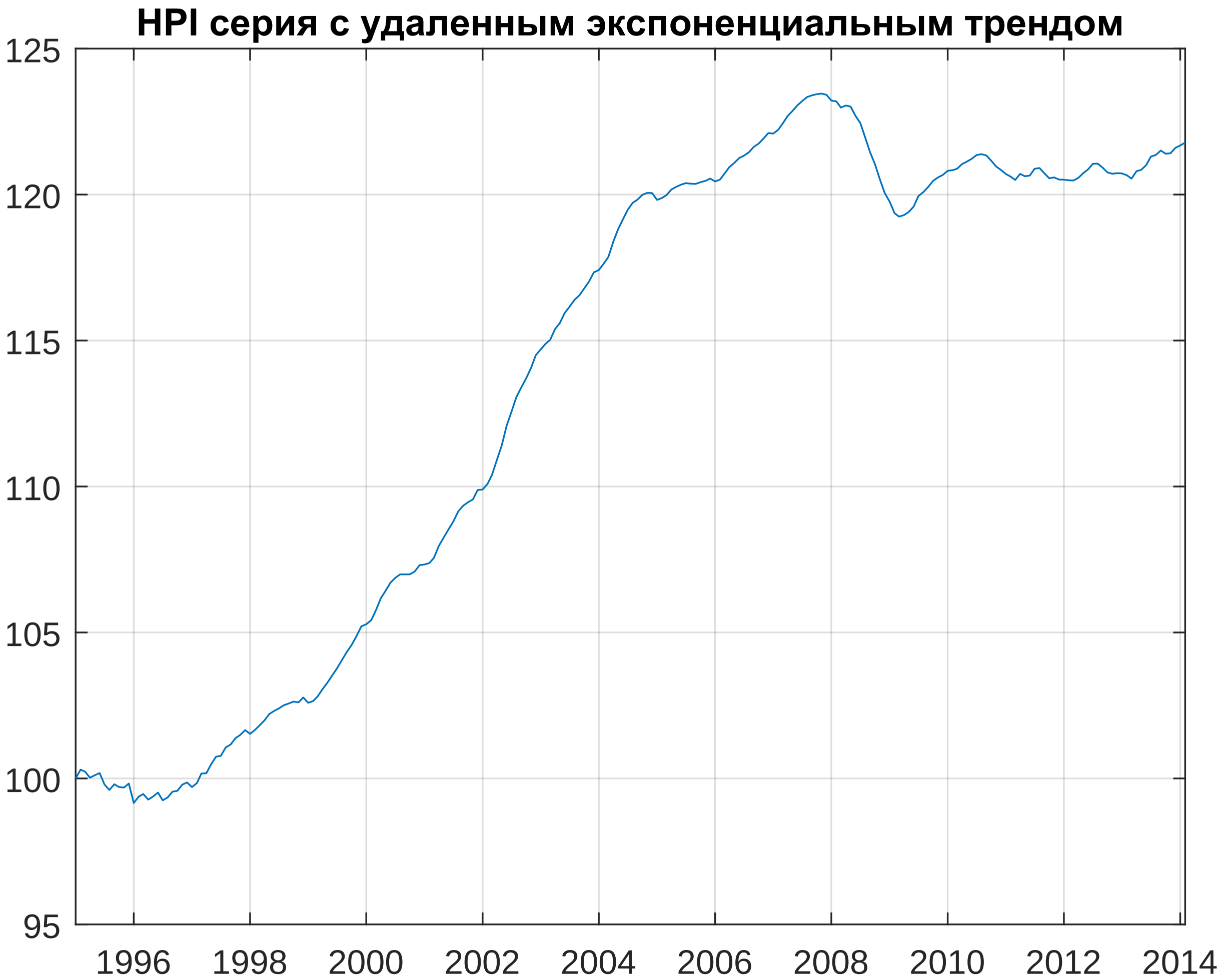


Рис. 1.5: Выявление экпоненциального тренда

Взятие разности

```
first_diff = LagOp([1, -1], 'Lags', [0,1]); % Задаем полином лагового оператора
dHPI = filter(first_diff, HPI_log); % Применяем лаговый оператор
lagdates = dates(2:end); % Временная переменная после взятия разности

plot(lagdates, dHPI)
title('Первая разность от DS-ряда "HPI\_log"')
grid on
```

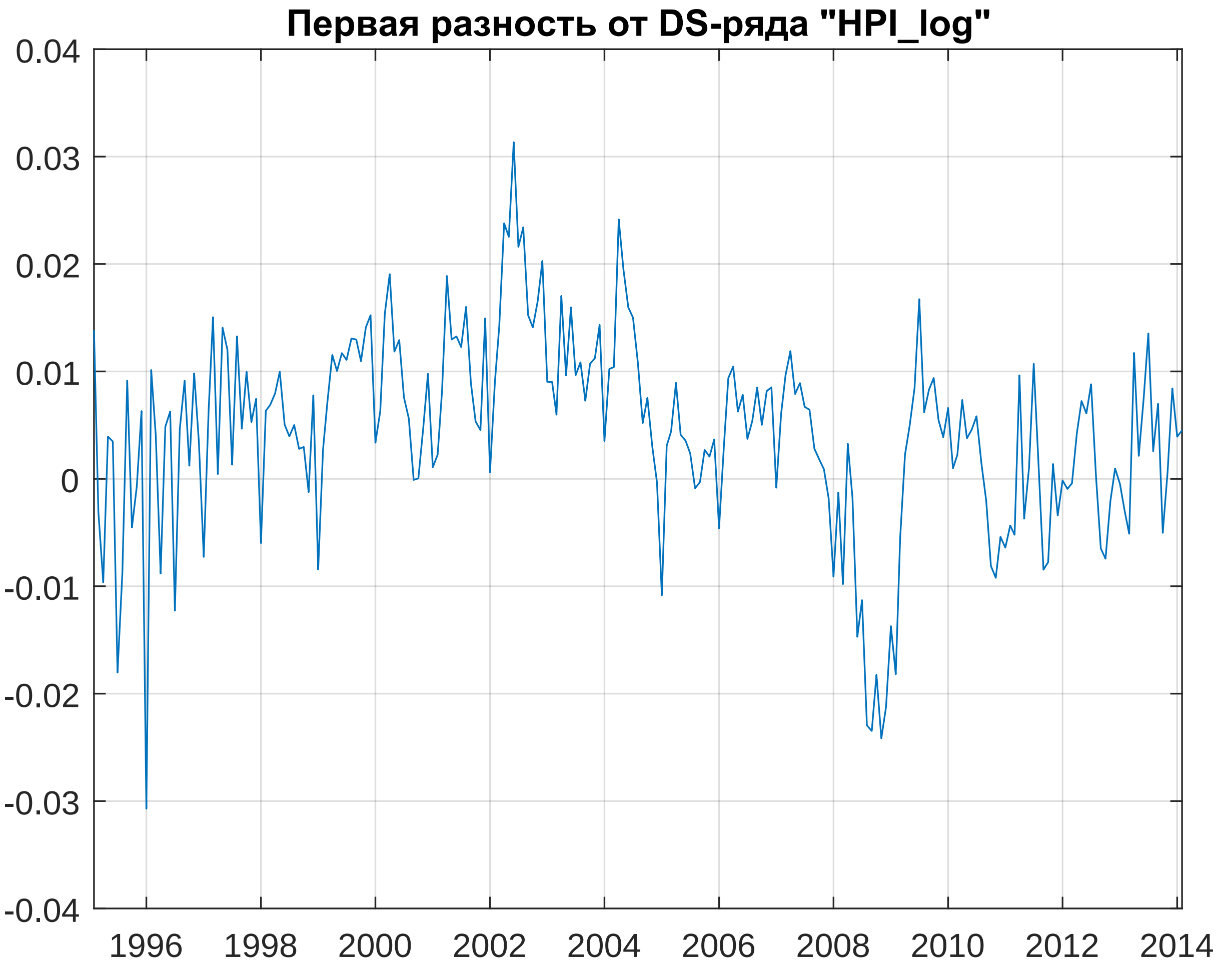


Рис. 1.6: Разность от интегрированного ряда

Тест на стационарность

```
n_test_lags = floor(sqrt(numel(dHPI)));
h = kpsstest(dHPI, 'lags', n_test_lags, 'trend', true);
disp('Проведен тест Квятковского-Филлипса-Шмидта-Шина')
```

```
Проведен тест Квятковского-Филлипса-Шмидта-Шина
```

```
disp(['Решением об отклонении нуль гипотезы: ',num2str(h)])
```

```
Решением об отклонении нуль гипотезы: 1
```

```
if h==1
    disp('Ряд нестационарен')
elseif h==0
    disp('Ряд стационарен')
end
```

```
Ряд нестационарен
```

Идентификация полиномиального тренда

```
dn = days(lagdates-lagdates(1)); % Отностельое время в днях
[trend_fit, ~, date_stats] = polyfit(dn, dHPI, 3); % Подбор кубического полинома
trend_val = polyval(trend_fit, dn, [], date_stats); % Значения полиномиального тренда
detrend = dHPI - trend_val; % Детрендовый ряд

plot(lagdates,detrend)
title('Детрендовый ряд')
xlabel('Даты')
grid on
```

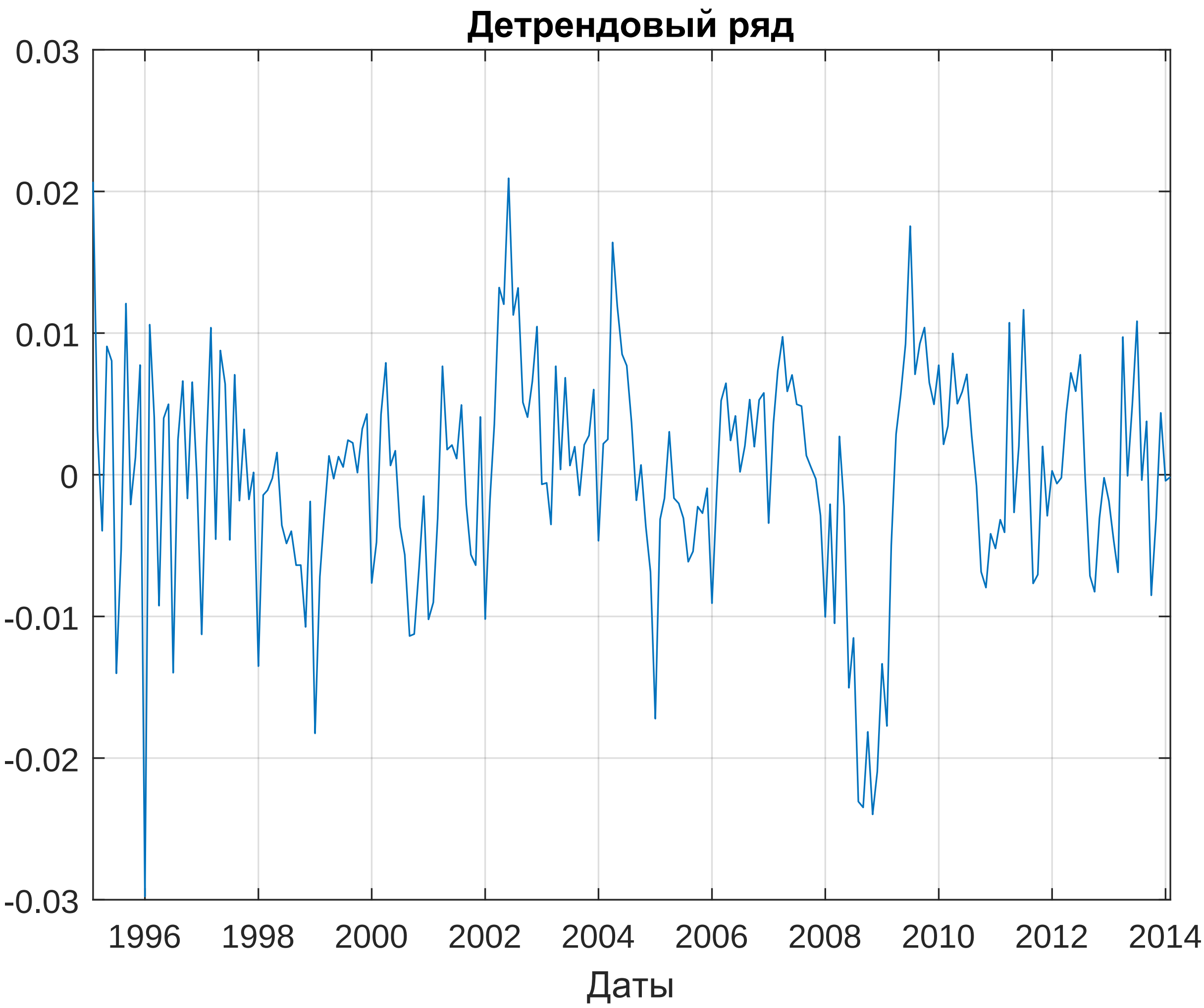


Рис. 1.7: Выявление полиномиального тренда

Тест на стационарность 2

```
n_test_lags = floor(sqrt(numel(detrend)));
h = kpsstest(detrend, 'lags', n_test_lags, 'trend', false);
disp('Проведен тест Квятковского-Филлипса-Шмидта-Шина')
```

```
Проведен тест Квятковского-Филлипса-Шмидта-Шина
```

```
disp(['Решением об отклонении нуль гипотезы: ',num2str(h)])
```

```
Решением об отклонении нуль гипотезы: 0
```

```
if h==1
    disp('Ряд нестационарен')
elseif h==0
    disp('Ряд стационарен')
end
```

```
Ряд стационарен
```

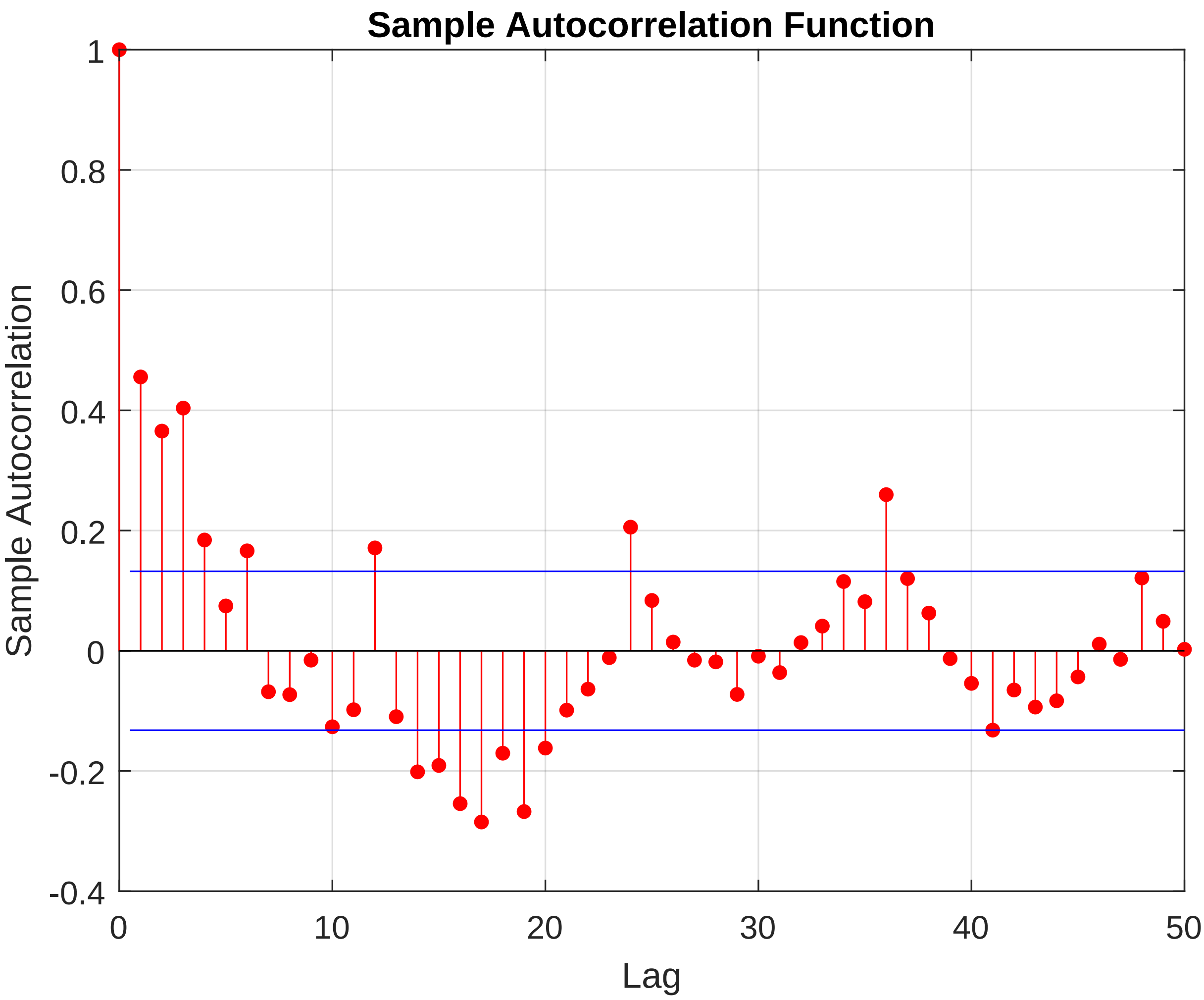


Рис. 1.8: Коррелограмма

На основании данных выборочной коррелограммы заключаем, что присутствует сезонная компонента. Включим её в будущую модель.

1.3.2 Выбор модели

Сравнение ACF и PACF

```
figure
nLag = 50;
subplot(2,1,1)
autocorr(statHPI, nLag)
subplot(2,1,2)
parcorr(statHPI, nLag)
```

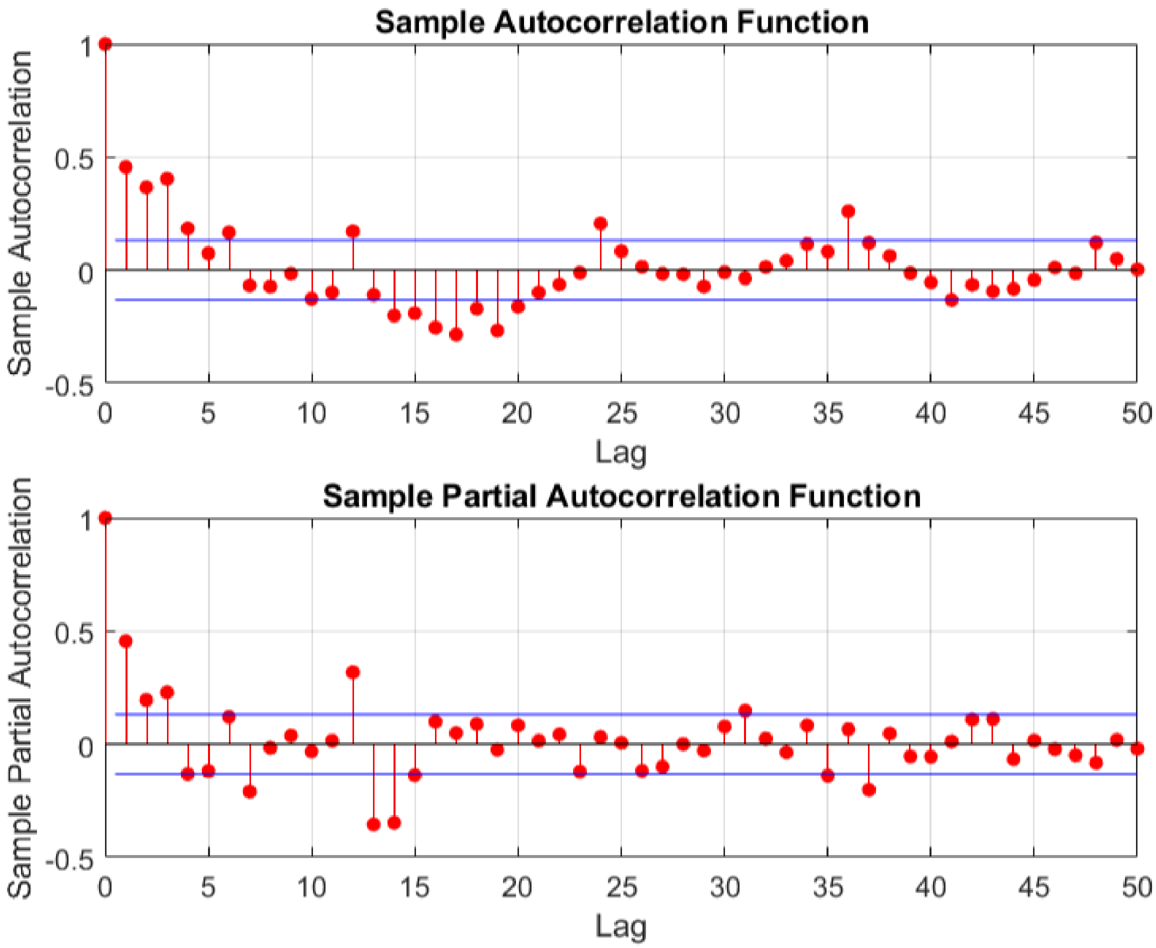


Рис. 1.9: Обычная и частная коррелограммы

Исходя из данных ACF и PACF коэффициенты p и q для модели ARMA(p,q) вычленить не удалось.

Идентификация моделей кандидатов

$$ARIMA(i, 0, j), \text{ где } i = 1 \dots 5, j = 1 \dots 5;$$

```
Mdl = cell(1,1);
for i =1:5
    for j = 1:5
        hpi_model = arima('D',0,...
            'Seasonality',12,...
            'ARLags', 1:i,...
            'MALags', 1:j,...
            'SARLags', 12, ...
            'Constant', 0);
        hpi_model = estimate(hpi_model,statHPI);
        Mdl{i,j} = summarize(hpi_model);
    end
end
```

Сравнение информационного критерия Акаике:

```
z = zeros(size(Mdl));
for i = 1:size(Mdl,1)
    for j = 1:size(Mdl,2)
        z(i,j) = Mdl{i,j}.AIC;
    end
end
figure
heatmap(z) % Карта значений AIC
```

```
title('Иформационного критерий Акаике в зависимости от (p, q)')
xlabel('p')
ylabel('q')
```

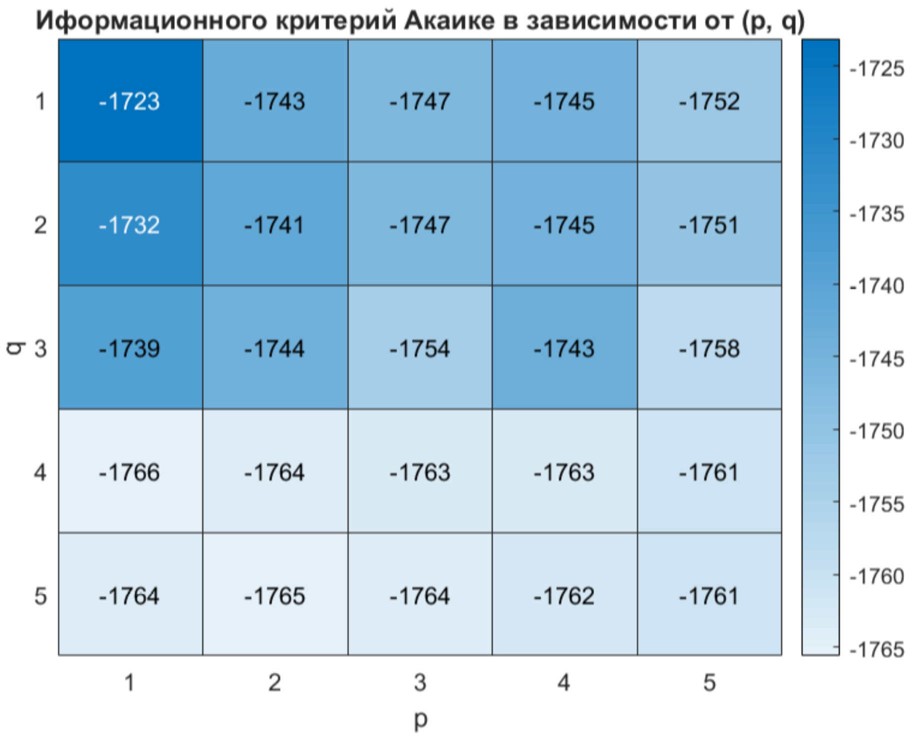


Рис. 1.10: Сравнение информационного критерия

Наименьшее значение AIC показывает модель $ARIMA(1, 0, 4)$.

Создание ARIMA модели и оценка ее коэффициентов

```
chosenModel = arima('D',0,...
                'Seasonality',12,...
                'ARLags', 1,...
                'MALags', 1:4,...
                'SARLags', 12, ...
                'Constant', 0);
[chosenModel, Q1] = estimate(chosenModel,statHPI); % Оценка коэффициентов модели
```

```
ARIMA(1,0,4) Model Seasonally Integrated with Seasonal AR(12) (Gaussian Distribution):
```

	Value	StandardError	TStatistic	PValue
Constant	0	0	NaN	NaN
AR{1}	0.82992	0.059376	13.977	2.1468e-44
SAR{12}	-0.34812	0.050162	-6.94	3.9224e-12
MA{1}	-0.49466	0.080163	-6.1707	6.8002e-10
MA{2}	0.1969	0.068953	2.8556	0.0042951
MA{3}	0.18671	0.071682	2.6047	0.0091964
MA{4}	0.021114	0.08247	0.25602	0.79793
Variance	2.7053e-05	2.2199e-06	12.187	3.6582e-34

Замечание: Учитывая p-value для оценки коэффициента при MA члене 4-го порядка, удалим его из модели, в силу его незначимости на 5%-уровне.

1.3.3 Оценка адекватности модели

Анализ остаточного распределения

```
res = infer(chosenModel, statHPI); % Остатки в модели

figure
plot(lagdates, res)
title('Прогнозная серия остатков для модели ARIMA с постоянной дисперсией')
grid on
```

(См. рис. 1.11)

Среднее значение остатков: $-5.7221 \cdot 10^5$.

Визуализация распределения остатков:

```
figure
subplot(1, 2, 1)
histogram(res, 20)
xlabel('Data')
title('Гистограмма остатков')
subplot(1, 2, 2)
normplot(res)
title('Нормальная вероятностная бумага')
```

(См. рис. 1.12)

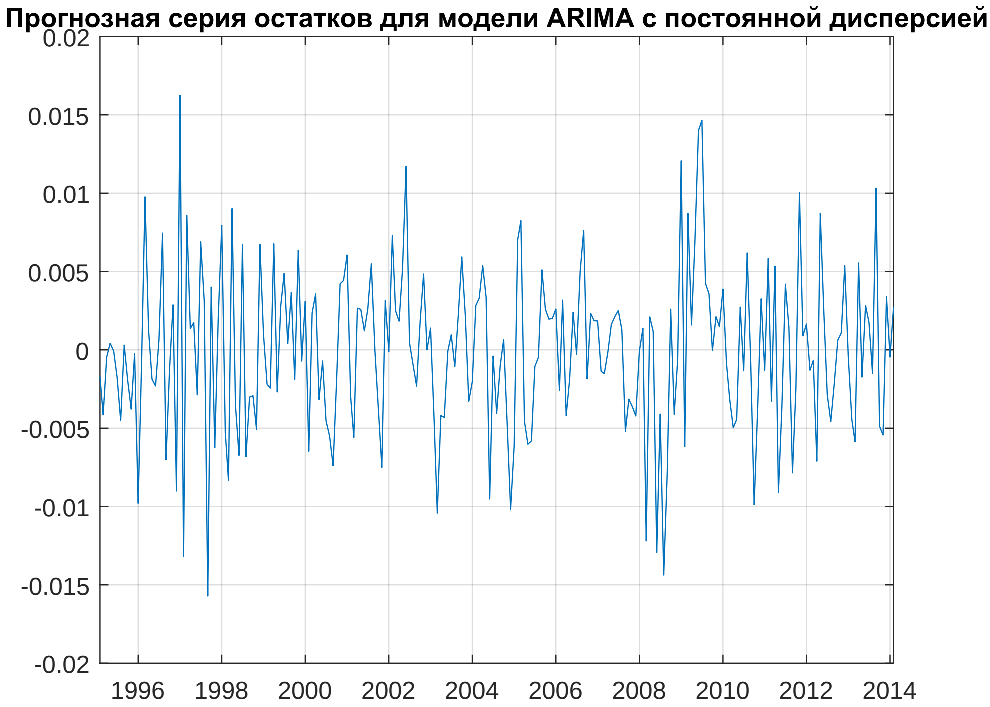


Рис. 1.11: Серия остатков

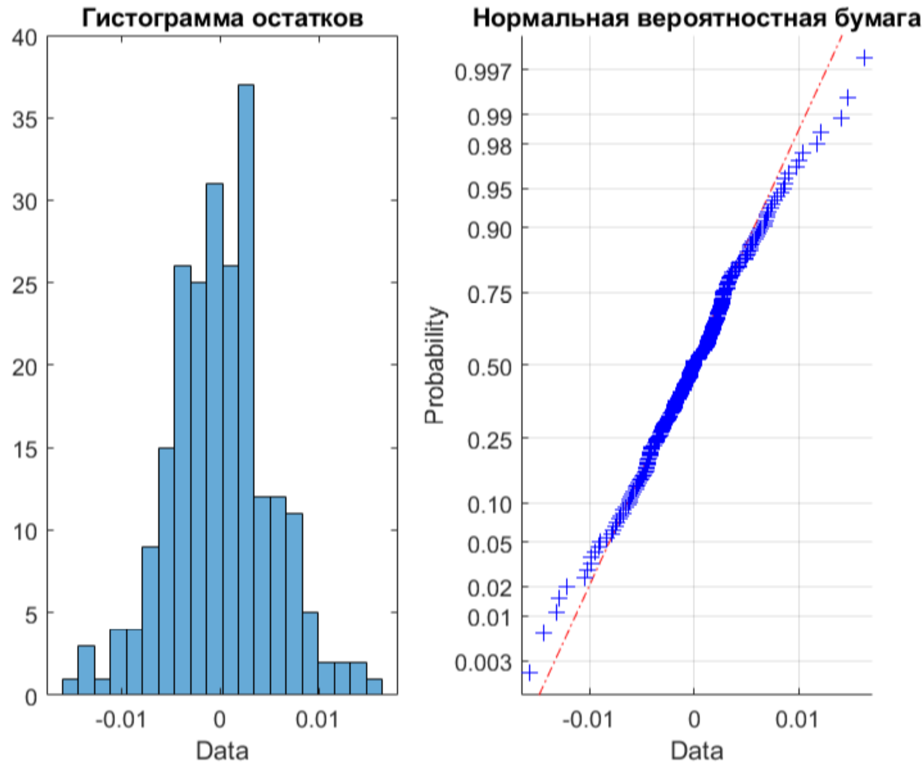


Рис. 1.12: Распределение остатков

Отображение результатов формального тестирования:

```
T = table(~jbtest(res), ~adtest(res), ~lillietest(res), ...
    'VariableNames', {'JB_NORMAL', 'AD_NORMAL', 'LILLIE_NORMAL'});
disp('Проверка нормальной гипотезы на остатках:')
```

```
Проверка нормальной гипотезы на остатках:
```

```
disp(T)
```

JB_NORMAL	AD_NORMAL	LILLIE_NORMAL
true	true	true

Анализ автокорреляции и дисперсии остатков

Количество оцениваемых коэффициентов

```
SAR_coeffs = cell2mat(chosenModel.SAR);
n_est_coeffs = numel(chosenModel.AR)+numel(chosenModel.MA)+nnz(SAR_coeffs);
```

Тест гипотезы на автокорреляцию остатков.
(Нуль гипотеза: отсутствие автокорреляции)

```
Box_lags = min(20, numel(res)-1);
h = lbqtest(res, 'lags', Box_lags, 'dof', Box_lags-n_est_coeffs);
disp(['Результат Q-теста Льюнга-Бокса: ',num2str(h)])
```

```
Результат Q-теста Льюнга-Бокса: 1
```

Исследование условной гетероскедастичности

```
h = lbqtest(res.^2, 'lags', 1);
disp(['Q-тест Льюнга-Бокса на автокорреляцию квадратов ошибок: ',num2str(h)])
```

```
Q-тест Льюнга-Бокса на автокорреляцию квадратов ошибок: 1
```

Выполнение ARCH теста Энгла на условную гетероскедастичности.
(Нулевая гипотеза: отсутствие ARCH-эффектов)

```
h = archtest(res, 'lags', 1:5);
disp('ARCH тест Энгла на условную гетероскедантичность:')
```

```
ARCH тест Энгла на условную гетероскедантичность:
```

```
disp(h)
```

```
   1   1   1   1   1
```

Таким образом, основываясь на результатах тестов, мы можем заключить, что ряд остатков проявляет условную гетероскедастичность. Отразим данное поведение ряда в нашей модели.

Модель условной гетероскедастичности

Инициализация кандидатов моделей:

```
var_model{1} = garch(1, 1);
var_model{2} = garch(0, 1);
var_model{3} = garch(1, 2);
```

Оценка каждой модели и вывод статистики:

```
for k = numel(var_model):-1:1
    [var_model{k}, cov_mat{k}, logLik(k)] = estimate(var_model{k}, res);
end
```

```
Warning: Lower bound constraints are active; standard errors may be inaccurate.

    GARCH(1,1) Conditional Variance Model (Gaussian Distribution):

                  Value       StandardError    TStatistic      PValue
                _________     _____________    __________    __________

    Constant    1.3529e-06     1.4524e-06       0.93143         0.35163
    GARCH{1}           0.9       0.034414        26.152      9.4005e-151
    ARCH{1}           0.05        0.0553        0.90416         0.36591

    GARCH(0,1) Conditional Variance Model (Gaussian Distribution):

                  Value       StandardError    TStatistic      PValue
                _________     _____________    __________    __________

    Constant    2.2902e-05     2.3816e-06        9.6159      6.8478e-22
    ARCH{1}        0.14858       0.085268        1.7425        0.081423

    GARCH(1,1) Conditional Variance Model (Gaussian Distribution):

                  Value       StandardError    TStatistic      PValue
                _________     _____________    __________    _________

    Constant    1.0611e-05     3.9809e-06        2.6655      0.0076882
    GARCH{1}       0.43601         0.2273        1.9182       0.055084
    ARCH{1}        0.17259       0.095768        1.8022       0.071518
```

```
for k = numel(var_model):-1:1
    var_model_info{k} = summarize(var_model{k});
end
```

Сравнение моделей с использованием информационного критерия:

```
nEstParams = [var_model_info{1}.NumEstimatedParameters, ...
              var_model_info{2}.NumEstimatedParameters];
[aic, bic] = aicbic(logLik(1:2), nEstParams, numel(res));
```

Отчеты:

```
[~, posAIC] = min(aic);
[~, posBIC] = min(bic);
disp(['AIC предлагает модель #:', num2str(posAIC)])
```

```
AIC предлагает модель #:2
```

```
disp(['BIC предлагает модель #:', num2str(posBIC)])
```

```
BIC предлагает модель #:2
```

Оценка комбинированной ARIMA/GARCH модели

```
overallModel = arima('D', 0, 'Seasonality', 12, ...
                    'ARLags', 1, ...
                    'MALags', 1:4, ...
                    'SARLags', 12, ...
                    'Constant', 0, ...
                    'Variance', garch(0, 1));
```

Подбор параметров модели:

```
[overallModel, Q] = estimate(overallModel, statHPI);
```

Диагностика:

```
ARIMA(1,0,3) Model Seasonally Integrated with Seasonal AR(12) (Gaussian Distribution)

 Effective Sample Size: 229
 Number of Estimated Parameters: 7
 LogLikelihood: 882.763
 AIC: -1751.53
 BIC: -1727.49

             Value    StandardError    TStatistic    PValue
             _____    _____________    __________    ______

 Constant     0.00        0.00             NaN         NaN
 AR{1}        0.80        0.05           16.35        0.00
 SAR{12}     -0.30        0.06           -5.24        0.00
 MA{1}       -0.45        0.09           -5.11        0.00
 MA{2}        0.16        0.07            2.22        0.03
 MA{3}        0.22        0.07            2.91        0.00


GARCH(0,1) Conditional Variance Model (Gaussian Distribution)

             Value    StandardError    TStatistic    PValue
             _____    _____________    __________    ______

 Constant    0.00         0.00            8.95        0.00
 ARCH{1}     0.18         0.10            1.83        0.07
```

В итоге, приходим к $ARMA(1,3)$ модели с сезонной компонентой, дисперсия которой моделируется с помощью $ARCH(1)$.

1.3.4 Прогноз и симуляция

Прогноз

Создание прогноза:

```
n_forecat_period = 10; % Количество временных шагов для прогноза
[stat_forecast, forecast_mse, forecast_var] = ...
    forecast(overallModel, n_forecat_period, 'Y0', statHPI);
```

Вычисление прогнозных дат:

```
forecast_dates = lagdates(end) + calmonths(1:n_forecat_period).';
new_dates = [lagdates(end); forecast_dates]; % Даты
```

Создание пользовательской визуализации:

```
% Значения прогнозов и доверительные интервалы
forecast_vals = [statHPI(end); stat_forecast];
upperConfBound = forecast_vals + 1.96*sqrt([0; forecast_mse]);
lowerConfBound = forecast_vals - 1.96*sqrt([0; forecast_mse]);

figure
subplot(2, 1, 1)
hold on
plot(lagdates, statHPI, 'b')
plot(new_dates, forecast_vals, 'k')
plot(new_dates, upperConfBound, 'r', 'LineStyle', ':')
plot(new_dates, lowerConfBound, 'r', 'LineStyle', ':')
title('Прогноз для стационарного ряда HPI')
xlabel('Даты')
ylabel(sprintf('Стационарный ряд HPI \n и прогнозные значения'))
grid on
hold off
subplot(2, 1, 2)
plot(lagdates(end-4:end), statHPI(end-4:end), 'b', 'LineWidth', 1.5)
hold on
plot(new_dates, forecast_vals, 'k', 'LineWidth',1.5)
plot(new_dates, upperConfBound, 'r', 'LineWidth', 1.5, 'LineStyle', ':')
plot(new_dates, lowerConfBound, 'r', 'LineWidth', 1.5, 'LineStyle', ':')
title('Прогнозный период для стац. ряда HPI')
ylabel(sprintf('Стационарный ряд HPI \n и прогнозные значения'))
legend({'Стационарные данные HPI','Предсказания','Верхняя доверительная граница',...
    'Нижняя доверительная граница'},'Location','eastoutside')
hold off
```

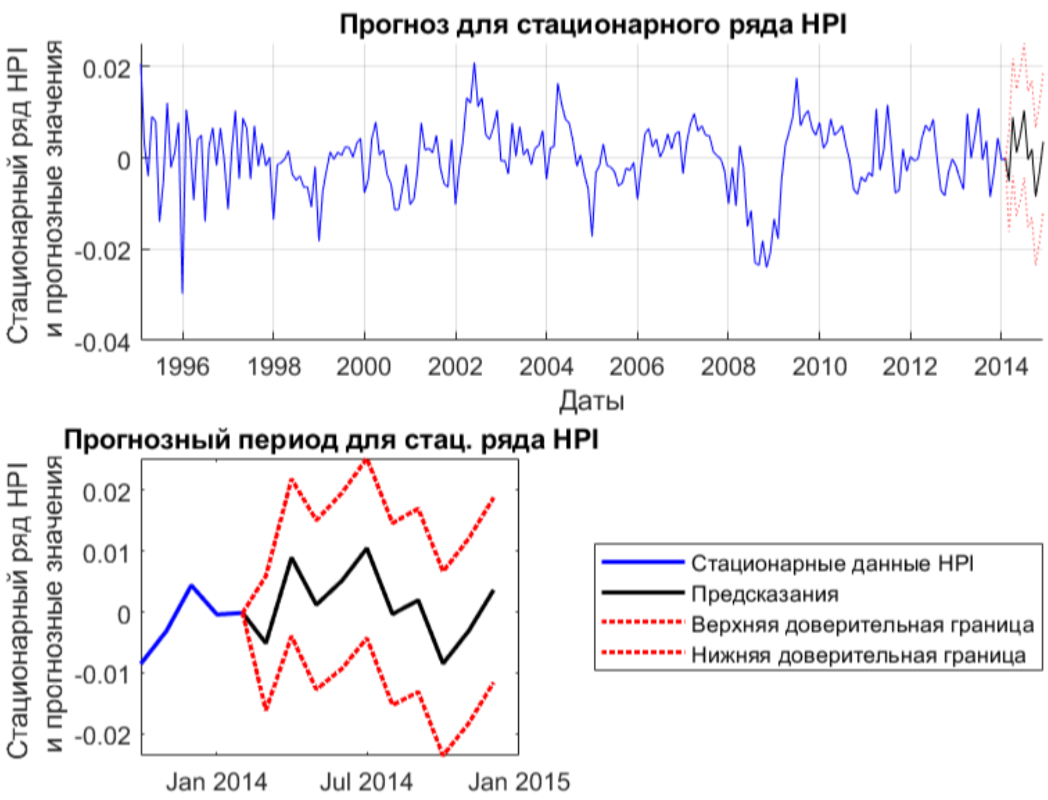


Рис. 1.13: Прогноз

Моделирование Монте-Карло

Расчет траекторий модели Монте-Карло для стационарных данных HPI:

```
n_forecat_period = 10;
statHPI_sim = simulate(overallModel, n_forecat_period, 'NumPaths', 1000, 'Y0', statHPI);
```

Извлекаем среднюю траекторию и ее доверительные интервалы:

```
required_prctiles = [2.5, 50, 97.5]; % Требуемые вероятности
n_prctiles = numel(required_prctiles);
m = (n_prctiles-1)/2;
statHPI_proj = prctile(statHPI_sim, required_prctiles, 2); % Процентили для прогноза

figure
hold on
plot(lagdates(end-4:end), statHPI(end-4:end), 'b', 'LineWidth', 1.5)
plot(new_dates, [repmat(statHPI(end), 1, size(statHPI_sim, 2)); statHPI_sim],...
    'HandleVisibility', 'off')
plot(new_dates, [statHPI(end); statHPI_proj(:, m+1)], 'k', 'LineWidth', 3)
plot(new_dates, [repmat(statHPI(end), 1, size(statHPI_proj, 2)-1);...
    statHPI_proj(:, [1:m, m+2:end])], 'r', 'LineWidth', 3, 'LineStyle','-')
title('Моделирование HPI с медианой и 95% доверительными интервалами')
xlabel('Даты')
ylabel(sprintf('Стац. ряд и траектории моделирования'))
legend({'стационарные данные HPI', 'медианная траектория',...
    'доверительные интервалы (2.5% и 97.5% процентов)'}, 'Location', 'southoutside')
```

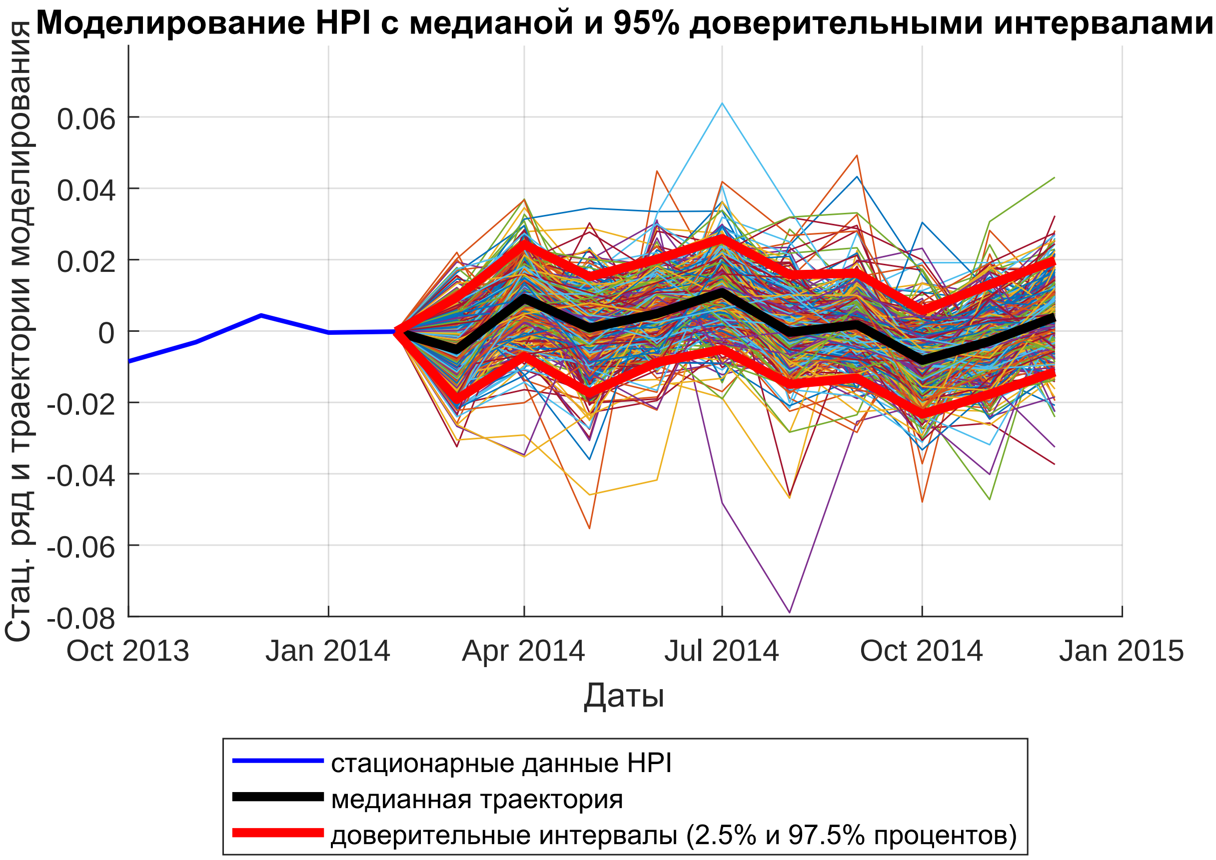


Рис. 1.14: Моделирование Монте Карло

Обратное преобразование

```
allDates = [lagdates; forecast_dates];
X = [repmat(statHPI, 1, 3); lowerConfBound(2:end), stat_forecast, upperConfBound(2:end)];
trend_to_add = polyval(trend_fit, days(allDates-dates(2)),...
    [], date_stats); % Генерация полиномиального тренда
X = X + repmat(trend_to_add, 1, 3); % Возвращаем полиномиальный тренд
X = cumsum([repmat(HPI_log(1), 1, 3); X]);%Коммулятивная сумма (обратна к взятию разности)
X = exp(X); % Возвращаем экспоненциальный тренд
```

Визуализация:

```
HPI = X(1:end-n_forecat_period, 2);
lower_HPI = X(end-n_forecat_period:end, 1);
forecast_HPI = X(end-n_forecat_period:end, 2);
upper_HPI = X(end-n_forecat_period:end, 3);

figure
hold on
plot(dates, HPI)
plot(new_dates, forecast_HPI)
plot(new_dates, upper_HPI, 'r', 'LineWidth',1.5, 'LineStyle',':')
plot(new_dates, lower_HPI, 'r', 'LineWidth',1.5, 'LineStyle',':')
title(sprintf('HPI с предсказанием и \n доверительными интервалами'))
xlabel('Даты')
ylabel('Значения HPI ряда и его прогноза')
legend({'Исходные данные HPI', 'Прогноз HPI', 'Верхний доверительный интервал',...
    'Нижний доверительный интервал'}, 'Location', 'eastoutside')
grid on
hold off
```

(См. рис. 1.15)

```
figure
plot(dates(end-4:end), HPI(end-4:end), 'b', 'LineWidth', 1.5)
hold on
plot(new_dates, forecast_HPI, 'k', 'LineWidth', 1.5)
plot(new_dates, upper_HPI, 'r', 'LineWidth', 1.5, 'LineStyle',':')
plot(new_dates, lower_HPI, 'r', 'LineWidth', 1.5, 'LineStyle',':')
title('Прогнозный период для HPI серии')
xlabel('Даты')
ylabel('Прогнозные значения')
legend({'Исходные данные HPI', 'Прогноз HPI', 'Верхний доверительный интервал',...
    'Нижний доверительный интервал'}, 'Location', 'eastoutside')
grid on
hold off
```

(См. рис. 1.16)

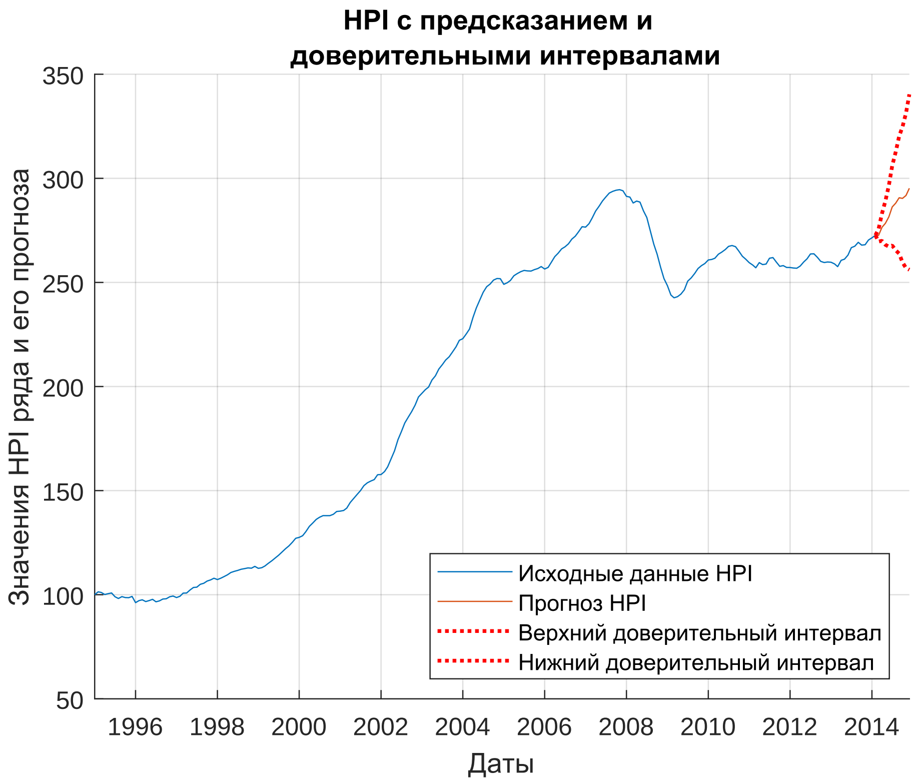


Рис. 1.15: Прогноз исходного временного ряда

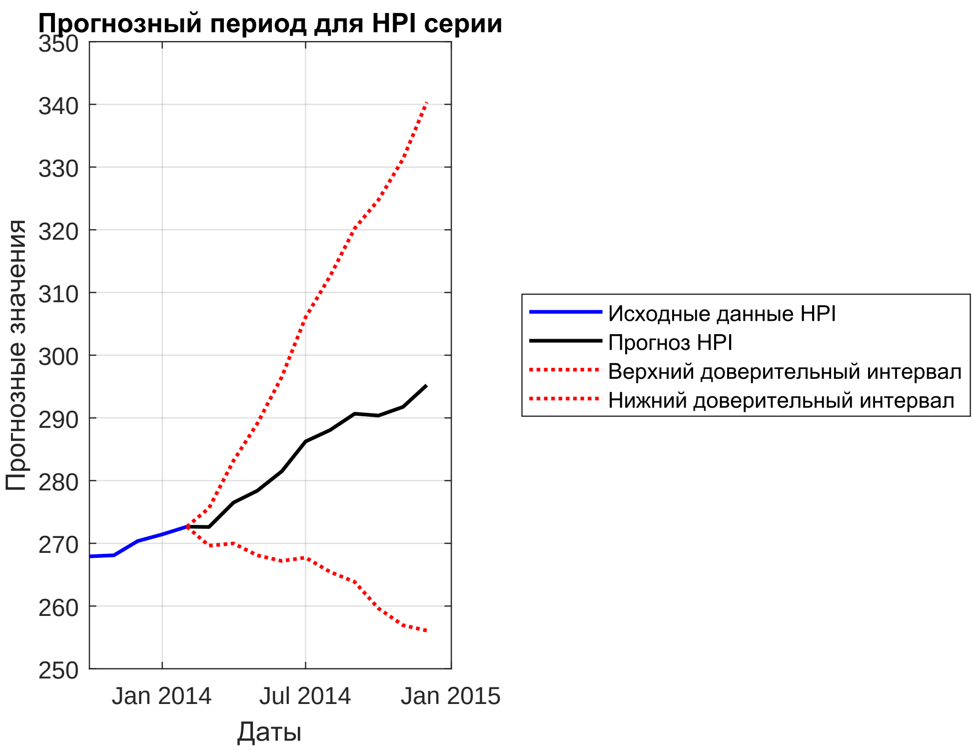


Рис. 1.16: Прогнозный период для исходного временного ряда

Глава 2

Задача 2: Взаимная корреляция временных рядов

2.1 Постановка задачи

Приведены ежедневные данные по заболеваемости и выздоровлению от COVID-19 в Москве в период 12.03.20-29.11.20 в формате xlsx. Вычислить среднее время выздоровления.

2.2 Решение поставленной задачи

2.2.1 Импортирование

```
T2 = readtable('time_series_COVID.xlsx','Sheet','Chart data', ...
    'Range','A2:C265', 'ReadVariableNames',false);
dates = T2.Var1; % Даты
infected = T2.Var2; % Заболевшие
recovered = T2.Var3; % Выздоровевшие
plot(dates, infected, dates, recovered)
title('Данные по COVID-19 в Москве, 15-29 ноября 2020г.')
xlabel('Даты')
ylabel('Кол-во человек')
legend('Заболеваемость','Выздоровление', 'Location','northeast')
grid on
```

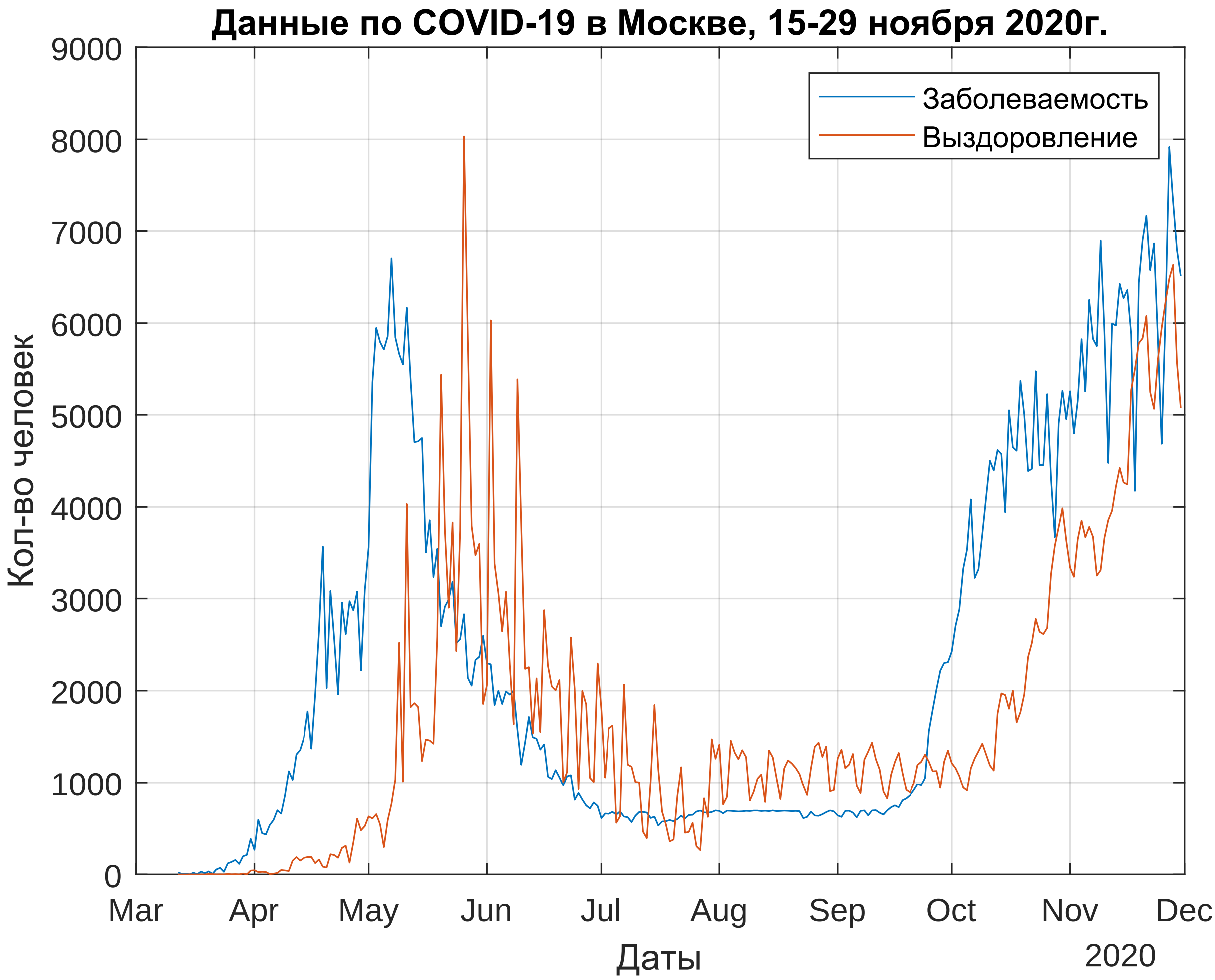


Рис. 2.1: Исходный временные ряды

2.2.2 Рассчет взаимной корреляции

```
rho = zeros(50,1); % Инициализация массива для корреляции
lags = 1:50; % Сдвиги
for k = lags
   rho(k) = corr(infected(1:end-k),recovered(k+1:end));
end
stem(lags,rho)
title('График взаимнокорреляционной функции')
xlabel('Лаги')
ylabel('Значения корреляции')
```

Значение лага, при котором наблюдается максимальная коррелированность, равно средней задержке между заболеванием и выздоровлением пациентов, что попросту говоря является временем их выздоровления.

```
[maxrho, recovtime] = max(rho);
disp("Среднее время выздоровления составляет " + recovtime +" дней.")
```

```
Среднее время выздоровления составляет 18 дней.
```

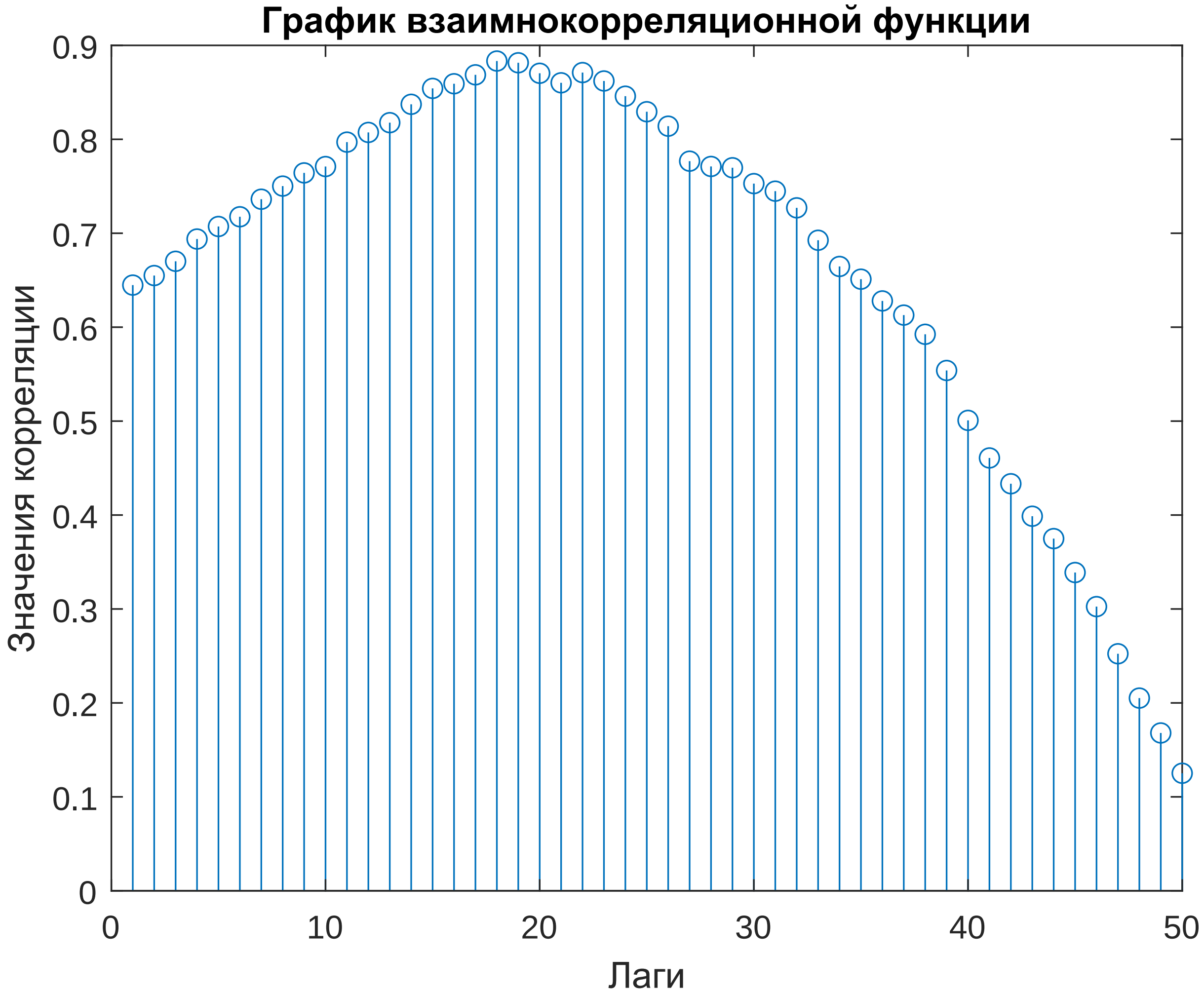


Рис. 2.2: Взаимная корреляция двух рядов

Заключение

При выполнении курсовой работы была поставлена и с помощью инструментария математического пакета MATLAB решена задача моделирования временного ряда представленного значениями индекса цен на недвижимость в Англии и Уэльсе за период с 1995г. по 2014г. Были изучены понятие, природа, примеры, типы временных рядов, а также методология их прогнозирования, на примере методологии Бокса-Дженкенса. Подробно разобраны многочисленные гипотез-тесты (такие как тесты на стационарность, гетероскедастичность, нормальность), их статистики и распределения. Проведена большая работа с данными с последующим сравнением моделей-кандидатов и оценкой коэффициентов в оптимальной модели, выбранной посредством минимизации информационного критерия Акаике. В результате расчетов получен прогноз поведения ряда на ближайшие 10 месяцев, а также доверительные границы этого прогноза с последующим тестированием методом Монте-Карло.

Во второй главе был сформулирован актуальный и остросоциальный вопрос о среднем времени выздоровления covid заболевших. Задача успешно решена путем рассчёта взаимной корреляции двух временных рядов с различными лагами.

Активно использованы статистические и эконометрические инструментарии и пакеты среды MATLAB. Большое внимание уделено работе с графикой. Результаты вычислений представлены в виде графиков и таблиц. В совокупности такой вывод результатов дает количественное и качественное представление о поведении временного ряда.

Все вычисления проводились на реальных данных, что позволяет сделать вывод о практической применимости, актуальности и востребованности описываемой теории.

Литература

[1] *Мишулина О. А.* Статистический анализ и обработка временных рядов. — М.: МИФИ, 2004. — С. 180.

[2] *Айвазян С.А.* Прикладная статистика. Основы эконометрики. Том 2. — М.: Юнити-Дана, 2001. — 432 с.

[3] *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс. — М.: Дело, 2007. — 504 с.

[4] Эконометрика. Учебник / Под ред. Елисеевой И.И. — 2-е изд. — М.: Финансы и статистика, 2006. — 576 с.

[5] *Носко В.П.* Эконометрика. Введение в регрессионный анализ временных рядов. — М., 2002. — 273 с.

[6] *Кремер Н.Ш.*, Путко Б.А. Эконометрика. — М.: Юнити-Дана, 2003-2004. — 311 с.

[7] *Dickey D. A. and Fuller W. A.* Distribution of the Estimators for Autoregressive Time Series with a Unit Root // Journal of the American Statistical Association. — 74. — 1979. — p. 427—431.

[8] ЦИТМ Экспонента [Электронный ресурс]: документация и семинары для пользователей MATLAB, Simulink/ авт. курса А. Багров. — М.: 2020. — Режим доступа: https://exponenta.ru, свободный — Загл. с экрана.

[9] MathWorks [Электронный ресурс]: Help Center, Documentation. — US: The MathWorks, Inc. — Режим доступа: https://www.mathworks.com/help/matlab/, свободный — Загл. с экрана.

[10] Онлайн-сервис для аналитики и визуализации бизнес-данных из различных источников Yandex DataLens [Электронный ресурс]: Коронавирус, дашборд. — Электрон. дан. — М.: 2020. — Режим доступа: https://datalens.yandex.ru/5v4ke2285sbk0?tab=X1, ограниченный — Загл. с экрана.