

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ**

Конкурс на лучшую работу по экономической тематике среди студентов
учреждений высшего образования

**Разработка и применение моделей MS-VARX для анализа
экономических циклов**

Макаревич Анатолий
Сергеевич,
студент 4 курса

Научный руководитель:
Малюгин Владимир
Ильич
доцент, кандидат
физико-математических
наук, кафедра ММАД

Минск, 2018

Abstract

This paper examines methods for extraction of economic cycles, identification of their turning points, and prediction of future values. This paper also compares the results of said methods when used to identify cycles in the GDP of the Republic of Belarus. Methods reviewed: double Hodrick-Prescott filtering, Hamilton's proposed method, Markov-switching autoregressive models (MS-ARX), and SARIMAX models. The work also includes a short description of the development of a Python-language library for time series analysis and modelling, which includes implementations of the above models.

Аннотация

В данной работе рассматриваются методы выделения экономических циклов, определения их поворотных точек, и предсказания будущих значений. Также сравниваются результаты этих методов при идентификации циклов в ВВП Республики Беларусь. Рассмотренные методы: двойное использование фильтра Ходрика - Прескотта, метод Хамильтона, авторегрессионные модели с Марковским переключением состояний (MS-ARX), модели SARIMAX. В работе также приводится краткое описание разрабатываемой библиотеки программ на языке Python, предназначенной для анализа и моделирования временных рядов. Библиотека включает разделы, посвященные вышеописанным моделям.

Реферат

Курсовая Работа: 37 страниц, 16 рисунков, 4 таблиц, 23 источников

Ключевые слова: ИНДЕКС ЭКОНОМИЧЕСКИХ НАСТРОЕНИЙ, АНАЛИЗ ПОВОРОТНЫХ ТОЧЕК ЭКОНОМИЧЕСКИХ ЦИКЛОВ, МЕТОД ХОДРИКА-ПРЕСКОТТА, МЕТОД ХАМИЛЬТОНА, МОДЕЛИ С ПЕРЕКЛЮЧЕНИЕМ СОСТОЯНИЙ, ОЦЕНКА ТОЧНОСТИ ПРОГНОЗОВ, БИБЛИОТЕКА НА ЯЗЫКЕ PYTHON

В работе представляются результаты, полученные автором при решении задач, связанных с построением и применением моделей с марковскими переключениями состояний из семейства MS-VARX и сезонной ARIMA-модели (SARIMAX), использующих индекс экономических настроений и индексы доверия белорусской экономики на основе опросных данных системы мониторинга Национального банка Республики Беларусь.

Исследования в данном направлении проводились в Белорусском государственном университете в 2016-2017 гг. в рамках НИР "Разработка системы опережающих экономических индикаторов и экономических диффузных индексов для основных видов экономической деятельности и экономики Республики Беларусь в целом с использованием экономико-математических, эконометрических методов и моделей на основе данных системы мониторинга предприятий Национального банка Республики Беларусь".

Модельный и программный инструментарий для построения указанных индексов и их применения в предиктивных эконометрических моделях для реального ВВП, а также в моделях с марковскими переключениями состояний для анализа бизнес-цикла белорусской экономики представлены в заключительном отчете о НИР.

Содержание

1	Проблема анализа экономических циклов и основные подходы к ее решению	4
1.1	Понятие экономического цикла и его интерпретация	4
1.2	Методы выделения циклической составляющей	4
1.3	Методы оценивания поворотных точек	5
2	Эконометрические модели и методы анализа циклических изменений в экономике	6
2.1	Методы выделения и сравнительного анализа экономических циклов	6
2.1.1	Метод выделения циклов, основанный на фильтре Ходрика – Прескотта	6
2.1.2	Метод сезонной корректировки и выделения циклов Хамильтона	6
2.1.3	Анализ поворотных точек экономических циклов на основе экономических индикаторов	7
2.2	Модели семейства MS-VARX и их применение для анализа циклических изменений	7
2.2.1	Модель MS-ARX	8
3	Экспериментальное исследование методов анализа циклических изменений в экономике	10
3.1	Исследование метода выделения циклов на основе фильтра Ходрика – Прескотта	10
3.2	Исследование метода Хамильтона и его сравнительный анализ с методом на основе фильтра Ходрика – Прескотта	11
3.3	Анализ циклических изменений и оценка поворотных точек экономических циклов на основе модели MS-AR	17
3.4	Сравнительный анализ и экономическая интерпретация поворотных точек на основе различных методов	19
4	Прогнозная способность моделей	22
4.1	Задача валидации модели	22
4.2	Сравнение точности прогнозов моделей для ВВП	23
5	Разработка библиотеки Time Series for Data Science	25
5.1	Краткое описание и основные задачи	25
5.2	Принципы разработки библиотеки	26
5.3	Особенности имплементации	27
5.4	Результаты разработки	28

Глава 1

Проблема анализа экономических циклов и основные подходы к ее решению

1.1 Понятие экономического цикла и его интерпретация

В рамках концепции экономического цикла, используемой в НБЭИ (Национальное бюро экономических исследований) США, подразумевается последовательная смена двух фаз, называемых периодами «роста» (growth) и «спада» (recessions) экономической активности. При этом поворотные точки соответствуют «пику» (максимальной точке роста) и «дну» (минимальной точке спада) экономического цикла [12]. В рамках концепции ОЭСР (Организации экономического сотрудничества и развития) допускается детализация основных фаз цикла относительно долгосрочного тренда с выделением периодов «роста» и «замедления» (выше линии тренда), а также – «спада» и «восстановления» (ниже линии тренда) [13]. Поворотные точки в данном случае соответствуют моментам начала замедления роста и начала восстановления после спада.

Одной из ключевых задач анализа и прогнозирования экономической активности является разработка систем раннего обнаружения смены фаз экономических циклов на основе специально разработанных экономических индикаторов. Для получения ранних сигналов о смене фаз и оценивания моментов смены фаз, называемых поворотными точками, в рамках указанных систем применяются так называемые опережающие экономические индикаторы (leading economic indicators) [14]. В рамках [22; 23] была разработана система опережающих индикаторов для Республики Беларусь, в том числе Индекс Экономических Настроений/Economic Sentiment Indicator (ИЭН/ESI). В настоящей работе используются временные ряды значений индекса экономических настроений, а также реального ВВП Республики Беларусь в месячном исчислении с мая 2005 г. по январь 2017 г., полученные в рамках указанной НИР.

1.2 Методы выделения циклической составляющей

Для периодизации экономических циклов необходимо выделить из временных рядов Реального ВВП и ИЭН циклические составляющие, на основании которых затем оцениваются поворотные точки циклов. На

опережающий характер ИЭН указывает тот факт, что поворотные точки его цикла предшествуют поворотным точкам цикла реально ВВП. В настоящее время, основным методом выделения догосрочной циклической составляющей являются статистические фильтры. Наиболее популярным в настоящее время является фильтр Ходрика – Прескотта[5; 13]. Данный фильтр используется в указанной выше НИР, а так же в рамках данной работы.

В недавней статье Дж. Хамильтона [8] указывается на проблемы, возникающие при использовании фильтра Ходрика - Прескотта, и предлагается свой метод выделения долгосрочного цикла.

В проводимом исследовании используются оба метода, т.е. метод Ходрика – Прескотта и метод Хамильтона для выделения долгосрочного цикла, и проводится сравнительный анализ поворотных точек циклов, получаемых с помощью данных фильтров, а так же экспертным путем. Подобное исследование ранее не проводилось на белорусских данных [19].

1.3 Методы оценивания поворотных точек

Для оценки поворотных точек существуют несколько широко используемых методов. Один из самых популярных – алгоритм Брай - Бошана [4], который состоит из несколько этапов (с минимальной длительностью цикла в 15 месяцев и отдельных фаз в 5 месяцев). Другой популярный метод – модели с переключающимися режимами, впервые популяризированный работой Хамильтона [7]. Результаты Хамильтона с MS-AR для бизнес-цикла США во многом сошлись с периодизацией NBER. Многие научные исследования [2; 3; 9] указывают на сопоставимость Марковских моделей и традиционных методов, однако на практике редко используется из-за сравнительной трудоемкости и непрозрачности. Существуют и другие подходы, которые так же не используются из-за сравнительной сложности вычислений.

В данной работе используются оба подхода для решения указанной задачи.

Глава 2

Эконометрические модели и методы анализа циклических изменений в экономике

2.1 Методы выделения и сравнительного анализа экономических циклов

2.1.1 Метод выделения циклов, основанный на фильтре Ходрика – Прескотта

Приводим краткое описание метода выделения цикла на основе двойного применения фильтра Ходрика - Прескотта. Смысл фильтрации - выделение колебаний в разных частотных диапазонах. Временной ряд рассматривается в представлении $x_t = \tau_t + c_t + \epsilon_t$, где τ_t - тренд, c_t - циклическая составляющая, ϵ_t - шум. Фильтр проводит минимизацию функции:

$$f(\tau) = \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} (\tau_{t+1} - 2\tau_t + \tau_{t-1})$$

по значениям τ_t . Параметр λ определяет длительность циклов.

Согласно методике рассмотренной в [22; 23], эмпирическим путем были получены значения λ соответствующие выделению трендовой составляющей ($\lambda = 42131.155$ для циклов длиной 90 месяцев) и сглаживанию циклической составляющей ($\lambda = 13.93$ для циклов длиной 12 месяцев). Если сравнить с методологией ОЭСР [13], то отличается значение первого параметра из-за различной продолжительности долговременных циклов ($\lambda = 133107.94$ для циклов длиной 120 месяцев).

Полная процедура выделения циклов состоит из нескольких этапов. Идет нормализация ряда, затем происходит сезонная корректировка с X13-ARIMA-SEATS [17], выделяется и исключается долговременный тренд первым фильтром, а потом выделяется и сохраняется циклическая составляющая вторым фильтром (для удаление шума/сглаживания). Далее идет оценка поворотных точек (в методологии ОЭСР [13] используется алгоритм Брай-Бошана).

2.1.2 Метод сезонной корректировки и выделения циклов Хамильтона

Сравнительно новый метод выделения циклов был предложен Хамильтоном [7]. Он утверждает, что для широкого класса нестационарных

процессов (до $I(d)$ или с полиномиальным трендом порядка не выше d), регрессия y_{t+h} на y_t, \dots, y_{t-d+1} имеет стационарные остатки, и является способом выделения тренда и циклов с периодичностью h (т.е. сезонность). Оценивается регрессия такого вида:

$$y_{t+h} = \hat{\alpha}_0 + \hat{\alpha}_1 y_t + \hat{\alpha}_2 y_{t-1} + \dots + \hat{\alpha}_d y_{t-d+1} + \hat{\eta}_t$$

где остатки модели $\hat{\eta}_t$ - исходный ряд с убраннным трендом. В статье предлагается брать $d = 4$ как универсальный подход (действительно, интегрированность или детерминированные тренды выше четвертого порядка крайне редки или экономически необусловлены). Такое значение d было взято и для описанной в данной работе имплементации в R и Python [19].

2.1.3 Анализ поворотных точек экономических циклов на основе экономических индикаторов

Для анализа циклов также можно использовать модели, основанные на экономических индикаторах (Economic Indicator Analysis - EIA), в том числе опережающих индикаторов (Compound/Composite Leading Indicator - CLI). Построение таких новых индикаторов на основе уже существующих описано в методологии ОЭСР [14] для их исследований.

Методология ОЭСР использовалась для построения Индекса Экономических Настроек в рамках НИР [23]. В рамках этой НИР было разработано программное обеспечение на языке R **ESIanalysis**, а также модельный инструментарий, включающий предиктивные модели для темпов роста реального ВВП, а также модели с Марковским переключением состояний, в которых используется в качестве экзогенной переменной построенный ИЭН РБ. Ниже приводится описание новых моделей с Марковским переключением состояний, полученные автором работы, появившиеся после указанной НИР.

2.2 Модели семейства MS-VARX и их применение для анализа циклических изменений

Модели с переключением состояний (Regime Switching / *RS* models) – подкласс моделей со структурными изменениями в которых скачкообразно меняются параметры модели. Каждый «режим» (состояние) в данном случае соответствует отдельному набору параметров, и номер режима l является отдельной переменной.

Если l описывается Марковским процессом (текущее состояние зависит только от предыдущего, с матрицей вероятностей перехода), то такую модель называют *марковской моделью переключения состояний* (Markov Switching / MS model). Подробнее эти модели описаны в [11], [21].

Модель $MS(L) - VARX(p)$ (Markov-Switching Vector Autoregressive model with exogenous variables) описывается следующим уравнением:

$$y_t = \sum_{i=1}^p A_{i,l} y_{t-i} + B_l x_t + \eta_t$$

где L - количество классов/состояний, p - порядок авторегрессии, x_t - экзогенная векторная переменная, и η_t - нормально-распределенный белый шум с матрицей Σ_l .

$A_{i,l}$, $i = \overline{1, p}$ - авторегрессионные коэффициенты (A - матрица), B_l - матрица коэффициентов. Эти коэффициенты, как и Σ_l , определены для каждого режима l .

$l = l(t) \in \overline{1, L}$ - номер режима в момент времени t , с матрицей вероятности перехода:

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,L} \\ m_{2,1} & m_{2,2} & \dots & m_{2,L} \\ \dots & \dots & \dots & \dots \\ m_{L,1} & m_{L,2} & \dots & m_{L,L} \end{bmatrix}, \quad \sum_{i=1}^L m_{i,j} = 1 \quad \forall j \in \overline{1, L}$$

При $L = 2$, можно упростить матрицу:

$$M = \begin{bmatrix} \sigma_1 & 1 - \sigma_2 \\ 1 - \sigma_1 & \sigma_2 \end{bmatrix}$$

Оценивание параметров таких моделей можно проводить с помощью итерационного ЕМ-алгоритма (ЕМ MS-VARX) [11] либо методом максимального правдоподобия. Из-за стохастического характера алгоритма, сходимость к одному результату не гарантирована.

2.2.1 Модель MS-ARX

Если рассматривать одномерные временные ряды, то выделяется подкласс $MS(L) - AR(p)X$:

$$y_t = \sum_{i=1}^p \alpha_{i,l} y_{t-i} + \beta_l x_t + \eta_t$$

где η_t - одномерный гауссовский белый шум с матрицей σ_l . Параметры σ_l , $\alpha_{i,l}$ и β_l зависят от режима l .

Если брать одномерную экзогенную переменную с лагом k , то уравнение модели можно записать:

$$y_t = \sum_{i=1}^p \alpha_{i,l} y_{t-i} + \beta_l x_{t-k} + \eta_t$$

Обозначим эту модель $MS(L) - AR(p) - X(x[-k])$, . Оценивание можно проводить по вышеописанными алгоритмами для MS-VARX.

Глава 3

Экспериментальное исследование методов анализа циклических изменений в экономике

Следующая часть работы заключается в применение описанных выше методов на реальных данных описывающих экономику Республики Беларусь. Рассматриваются два ряда:

- Реальный ВВП (Real Gross Domestic Product/RGDP, в данной работе обозначение - GDP) Республики Беларусь, в качестве базового индикатора
- Индекс Экономических Настроений/ИЭН (Economic Sentiment Indicator/ESI) для РБ (из [18; 22; 23], в качестве опережающего индикатора.

Временные ряды Реального ВВП и ИЭН являются нестационарными (интегрированные порядка 1) с выраженной сезонностью. ВВП имеет ярковыраженный долговременный тренд. Для анализа циклических изменений в них сравнивались три метода.

Данные временные ряды представлены на рис. 3.1. Вертикальные линии - официальные оценки поворотных точек экономики, где каждая сплошная линия соответствует пику, а штриховая - дну.

3.1 Исследование метода выделения циклов на основе фильтра Ходрика – Прескотта

Рассматривались ряды, предварительно сезонно скорректированы процедурой X13-ARIMA-SEATS с помощью пакета Python **statsmodels** [15]. Используя вышеописанную процедуру двойной фильтрации, получились “гладкие” ряды (обозначены smoothed). Они представлены на рис. 3.2 и рис. 3.3. Во всех графиках, вертикальными чертами указывались поворотные точки ВВП, полученные экспертными оценками (сплошная соответствует пику, штрихованная - дну).

Как видно на рис. 3.5, полученный ряд ИЭН опережает ВВП на 4-5 периодов (месяцев), что подтверждает также кросс-корреляционная функция для сглаженных рядов (см. рис. 3.4). Самая сильная кросс-корреляция отмечается на этих лагах, слегка выше на четвертом. Это типичный результат для опережающих индикаторов. Также видно из рис. 3.5, что пики и локалии дна в сглаженном ВВП в целом соответствуют официальным поворотным

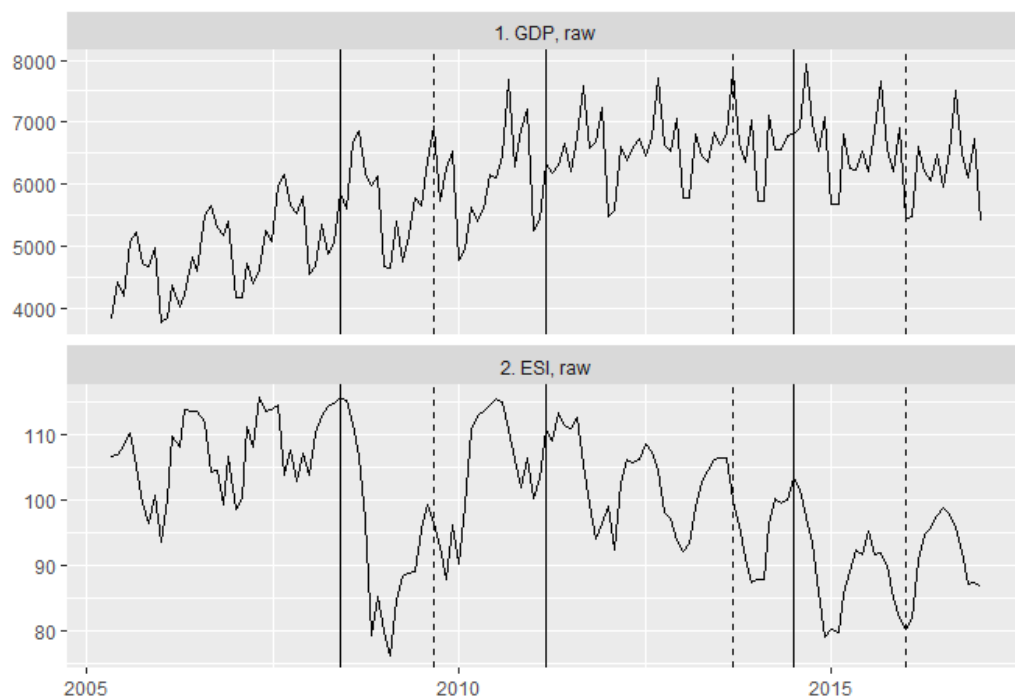


Рисунок 3.1 — Исходные ряды ВВП и ИЭН

точкам. Это дает уверенность в том, что этот метод выделения циклов работает для белорусской экономики.

3.2 Исследование метода Хамильтона и его сравнительный анализ с методом на основе фильтра Ходрика – Прескотта

В работе рассматривались и ряды, скорректированные по методу Хамильтона. Эта корректировка проводилась как с, так и без предварительной сезонной корректировки процедурой X13-ARIMA-SEATS [17]. Частичная регрессия (subset regression) вида $ARp(12, 13, 14, 15)$, которая требуется для корректировки по Хамильтону, проведена с помощью пакетов R **MSwM** и Python **statsmodels** [15]. Для определения поворотных точек ряды сглаживаются высокочастотным фильтром Ходрика-Прескотта ($\lambda = 13.93$). Результаты приведены на рис. 3.6, включая варианты с предварительной сезонной корректировкой. После корректировки по Хамильтону получаются похожие поворотные точки, но они могут отличаться на пару месяцев в разных направлениях. Ряды, скорректированные сезонно и по Хамильтону, имеют более схожие поворотные точки, с отклонениями в пару месяцев. Нужно также отметить, что первые $h = 12$ наблюдения в рядах, скорректированных по Хамильтону, невозможно использовать, поэтому для построения модели MS-ARX они исключены.

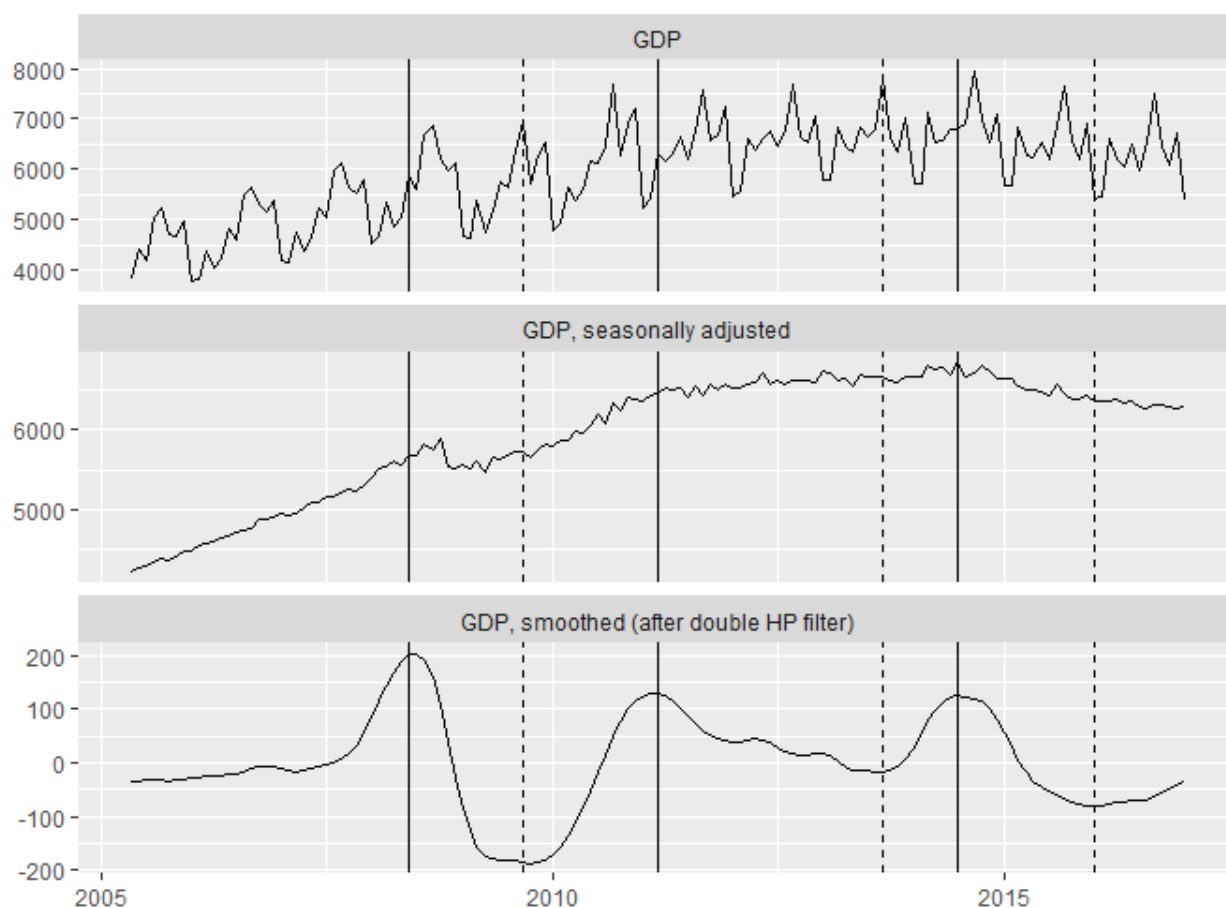


Рисунок 3.2 — ВВП, исходная серия (первый ряд), после сезонной корректировки (второй ряд), и после удаления тренда двойным применением фильтра Ходрика - Прескотта (нижний ряд).

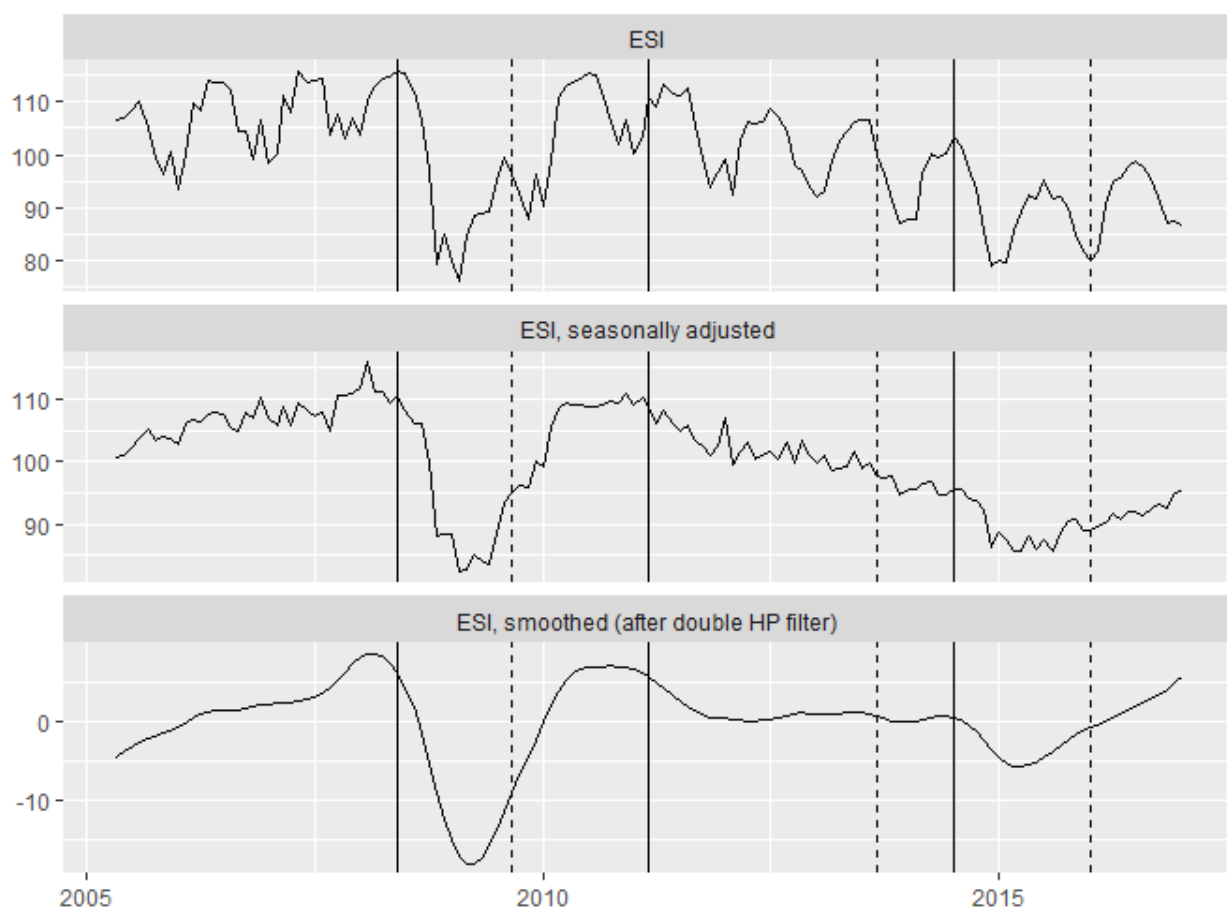


Рисунок 3.3 — ИЭН, исходная серия (первый ряд), после сезонной корректировки (второй ряд), и после двойного применения фильтра Ходрика - Прескотта (нижний ряд).

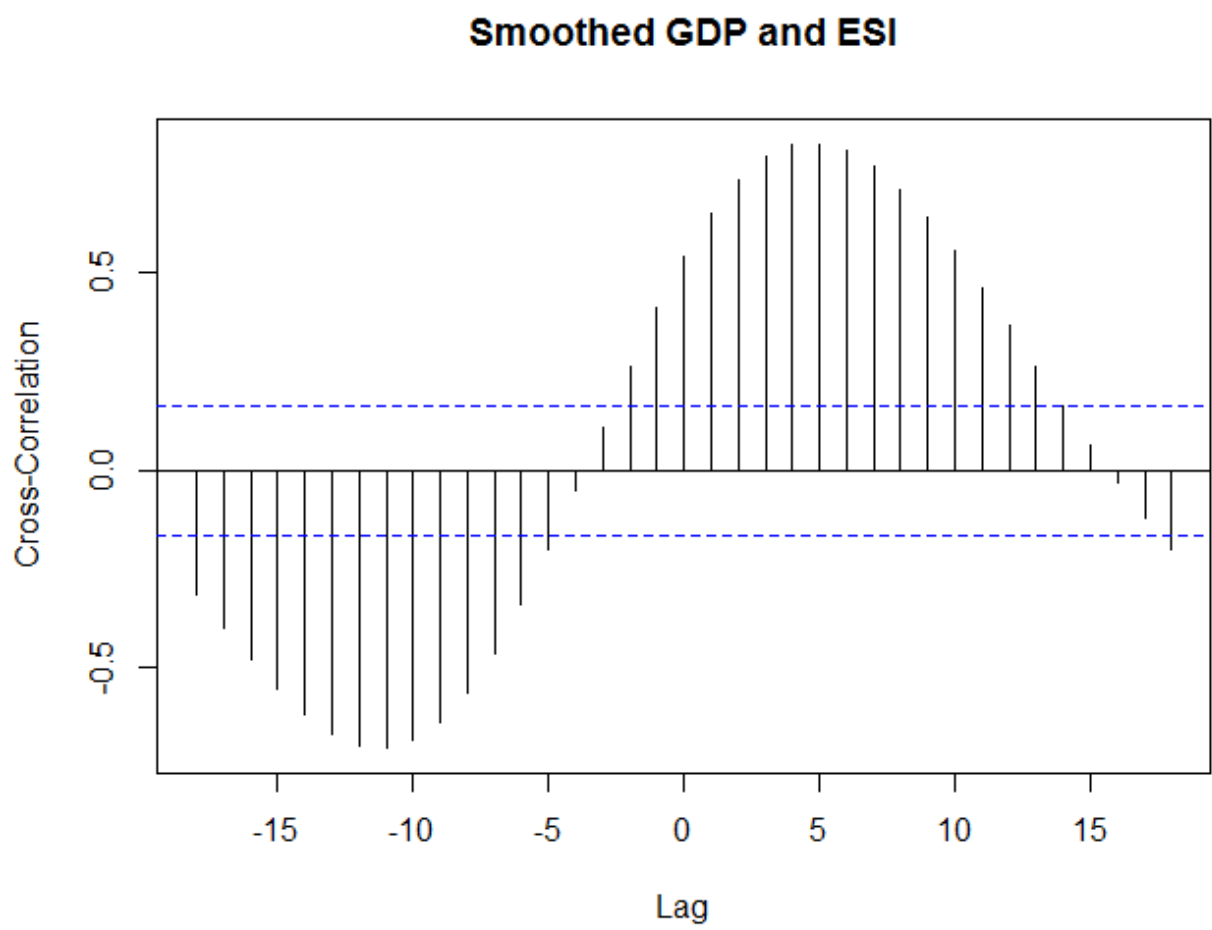


Рисунок 3.4 — Кросс-корреляция ВВП и ИЭН. Виден максимум на лагах 4-5.

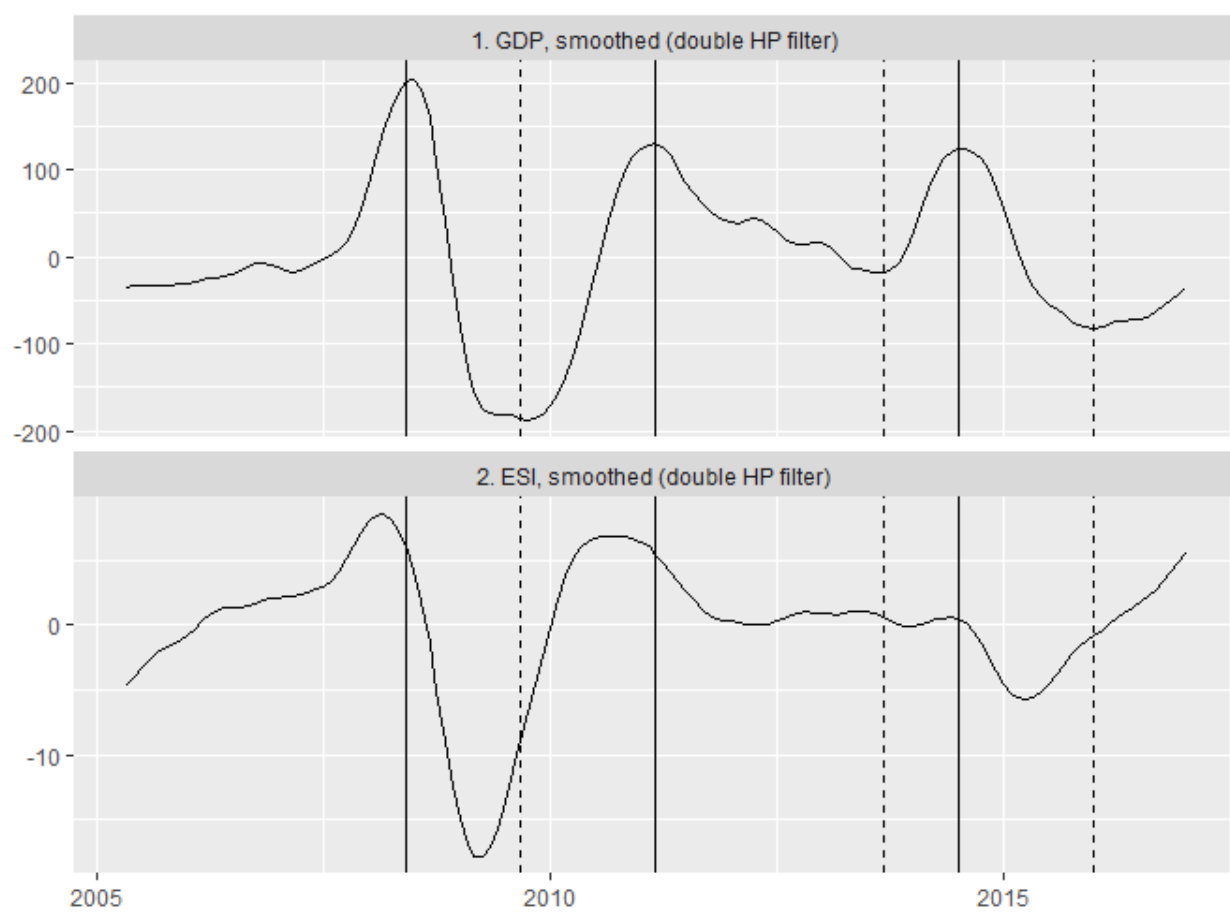


Рисунок 3.5 — Сравнение поворотных точек сглаженных ВВП и ИЭН.

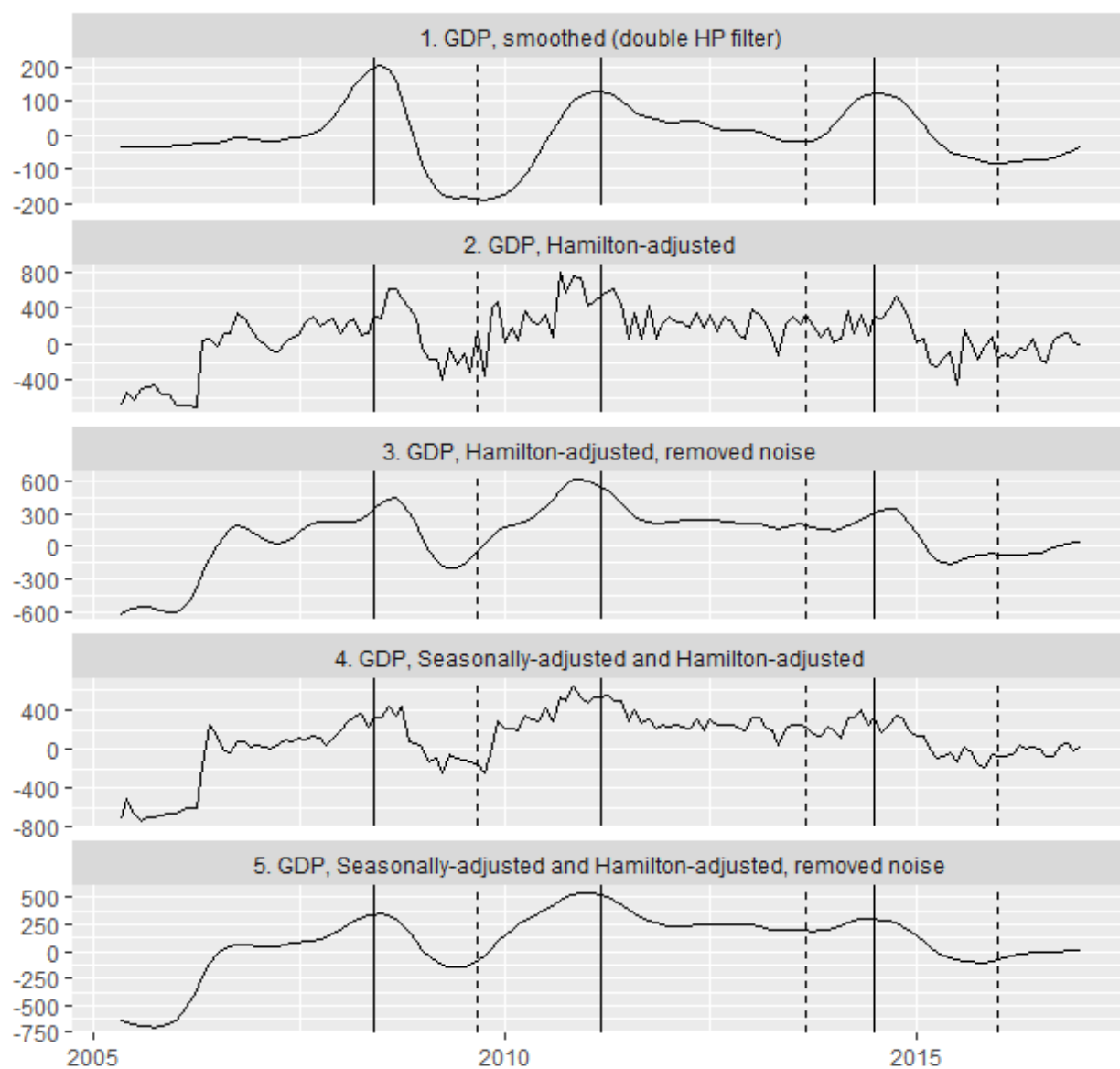


Рисунок 3.6 — Сравнение результатов методов выделения цикла на основании методов Ходрика - Прескотта и Хамильтона

В конце главы приведена таблица 3.3 сравнения поворотных точек GDP, полученные вышеописанными методами (а так же моделями MS-ARX, которые описаны в следующей части). Как видно, разные методы обработки временных рядов приводят к похожим, но не одинаковым оценкам времени поворотных точек (нагляднее всего заметить на рис. 3.6). Предлагается проводить сезонную корректировку до применения метода Хамильтона для получения значения поворотных точек, максимально близких к официальным [19; 20].

На основании приведенных в графиках и таблице результатов можно подтвердить применимость алгоритма Хамильтона для выделения циклов. Поворотные точки циклов ВВП и ИЭН, полученные на основе данного алгоритма, либо совпадают с экспертными оценками, либо отличаются на 1-2 месяца в сторону опережения. Главное несоответствие возникает в периоде 2011-2013 гг., который характеризуется высокой неопределенностью экономической конъюнктуры.

3.3 Анализ циклических изменений и оценка поворотных точек экономических циклов на основе модели MS-AR

Для временного ряда, скорректированного по Хамильтону, было построено несколько моделей $MS(L) - AR(p) - X(esi[-k])$ с параметрами $L \in [2, 3]$, $p \in [0, 3]$, $k \in [-1, 6]$. Для сравнения моделей использовались критерии:

- Значимость коэффициентов модели во всех режимах
- Критерий Акаике AIC
- Пороговое значение частоты переключения режимов (от 2 до 10 переключений). Это условие - проверка на адекватность режимов.

В результате была выбрана модель со спецификацией $MS(2) - AR(0) - X(esi[-4])$. Коэффициенты модели в табл. 3.1:

Таблица 3.1: Коэффициенты модели для ВВП по Хамильтону (и оценки значимости).

	<i>Value</i>	<i>Std.Err.</i>	<i>T - stat.</i>
$p[0- > 0]$	0.978684	0.015736	62.193815
$p[1- > 1]$	0.968666	0.024331	39.812009
$\alpha_{0,0}$	-0.204231	0.026092	-7.827207
$\alpha_{0,1}$	0.444673	0.025660	17.329314
β_0	-0.332577	0.090581	-3.671596
β_1	-0.233663	0.057333	-4.075552
σ_0^2	0.042897	0.007121	6.023806
σ_1^2	0.024106	0.005319	4.532125



Рисунок 3.7 — Предсказание модели MS(2)-ARX для ВВП по Хамильтону.

где $p[i- > i]$ - вероятность остаться в режиме i . Все коэффициенты оказались значимыми на уровне 0.05. Графики с режимами представлены на рис. 3.7, его остатки описаны на рис. 3.8.

Аналогичным методом была построена модель для годовых темпов роста ВВП и ИЭН¹. Была выбрана модель со спецификацией $MS(2) - AR(0) - X(esi[-2])$. Коэффициенты модели приведены в табл. 3.2.

Таблица 3.2: Коэффициенты модели для темпов роста (и оценки значимости).

	<i>Value</i>	<i>Std.Err.</i>	<i>T - stat.</i>
$p[0- > 0]$	0.979119	0.015488	63.218323
$p[1- > 1]$	0.971403	0.022110	43.935007
$\alpha_{0,0}$	-0.007623	0.003701	-2.059620
$\alpha_{0,1}$	0.097306	0.004468	21.778545
β_0	-0.143237	0.042976	-3.332932
β_1	-0.103041	0.036992	-2.785486
σ^2	0.000786	0.000100	7.875226

¹ от обоих рядов отнята единица, чтобы отцентрировать

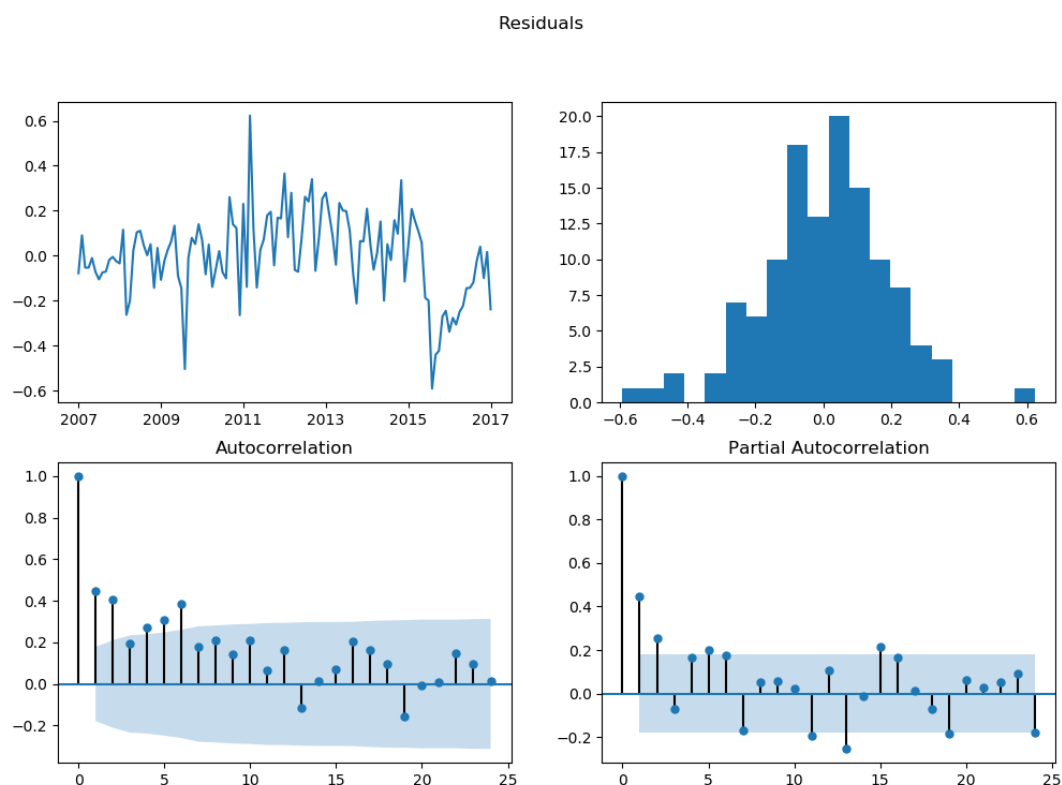


Рисунок 3.8 — Анализ остатков модели для ВВП по Хамильтону.

Графики ВВП, модельных значений, режимов, и анализа остатков так же приведены (рис. 3.9, 3.10).

3.4 Сравнительный анализ и экономическая интерпретация поворотных точек на основе различных методов

В таблице 3.3 приведены оценки поворотных точек разными методами. В качестве «истинных» значений рассматриваются экспертные оценки. Как видно, лучше всего совпадают точки, полученные в ходе двойной фильтрации (конкретно, для одной точки опаздывает на 1 месяц). Однако, точки полученные по Хамильтону (после сглаживания) отличаются не более чем на 2 месяца, причем всегда с опережением.

На основании приведенных в графиках и таблице результатов можно подтвердить применимость алгоритма Хамильтона для выделения циклов. Поворотные точки циклов ВВП и ИЭН, полученные на основе данного алгоритма, либо совпадают с экспертными оценками, либо отличаются на 1-2 месяца в сторону опережения. Главное не соответствие возникает в периоде 2011-2013 гг., который характеризуется высокой неопределенностью

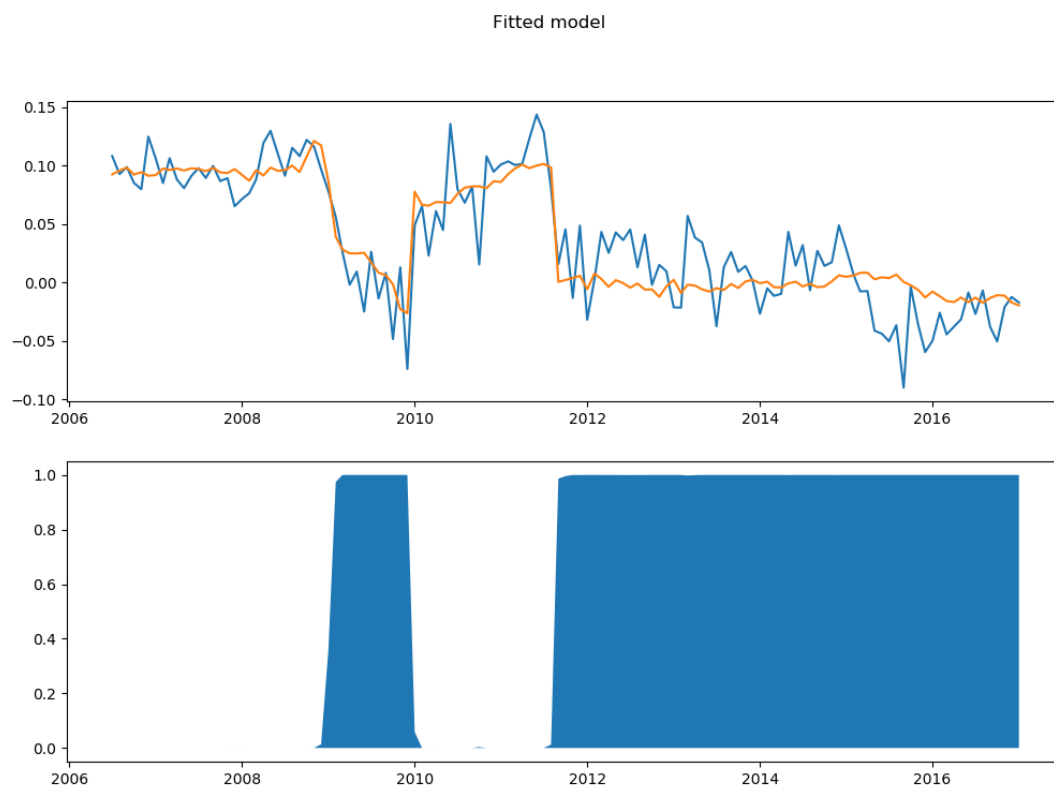


Рисунок 3.9 — Предсказание модели MS(2)-ARX для темпов роста ВВП.

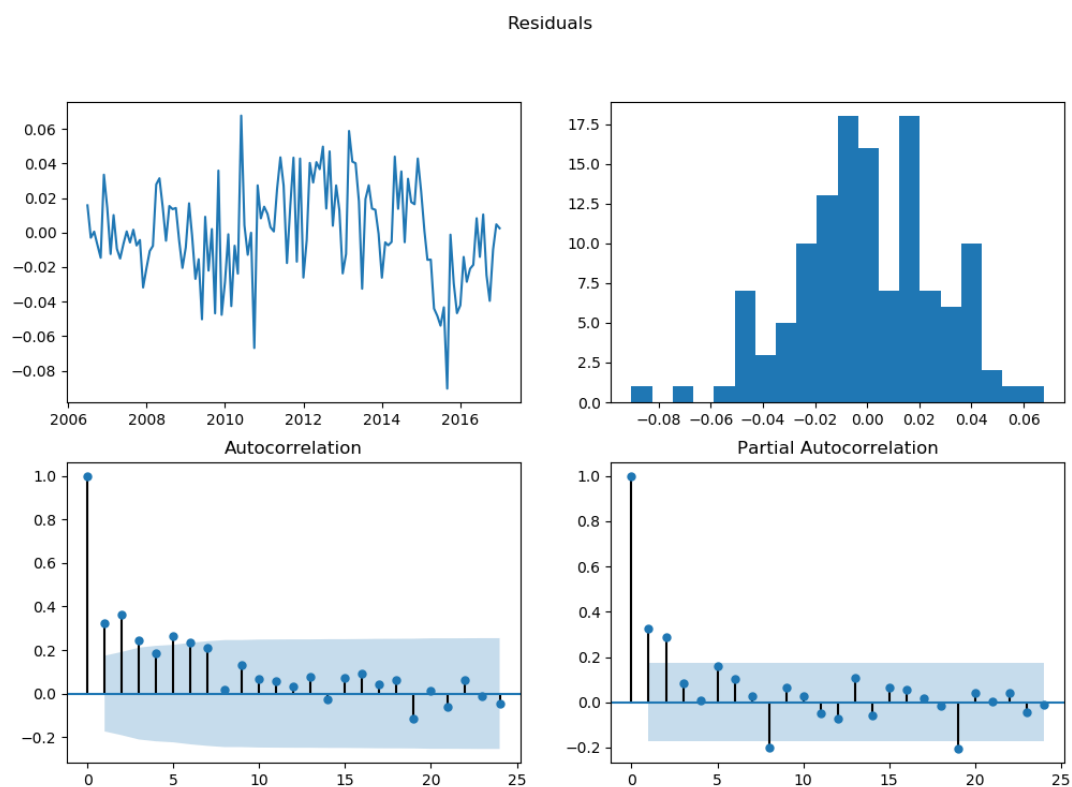


Рисунок 3.10 — Анализ остатков для модели темпов роста ВВП.

Таблица 3.3: Сравнение поворотных точек, полученные разными методами.

Метод	(пик 1)	(дно 1)	(пик 2)	(дно 2)
Экспертные оценки	2008.06	2009.09	2011.03	2013.09
Двойной HP filter	2008.06	2009.10	2011.03	2013.09
X13 + Hamilton (сглаж.)	2008.06	2009.07	2011.01	2013.09
MS-VAR для индикаторов доверия	2008.10	2009.02	2011.03	2012.06
MS-ARX для темпов роста ВВП (экзогенная - ИЭН)	2009.01	2010.01	2011.09	-
MS-ARX для Hamilton ВВП (экзогенная - ИЭН)	2009.01	2009.11	2011.08	-
Метод	(пик 3)	(дно 3)	Замечания	
Экспертные оценки	2014.07	2016.01	Официальные оценки	
Двойной HP filter	2014.07	2016.01	Методика совпадает с офиц.	
X13 + Hamilton (сглаж.)	2014.06	2015.11		
MS-VAR для индикаторов доверия	2014.11	-	см. [18]	
MS-ARX для темпов роста ВВП (экз. ИЭН)	-	-		
MS-ARX для Hamilton ВВП (экз. ИЭН)	-	-		

экономической конъюнктуры.

Точки переключения режимов моделей-подклассов MS-VARX для различных рядов сильнее отличаются от истинных поворотных точек. В сравнении с другими моделями, MS-ARX для рядов, скорректированных по Хамильтону более точно предсказала первые две точки, а также третий пик, но «пропустила» период 2011-2013 года. Есть предположение, что моделирование с помощью MS-VARX позволит этот период описать точнее. В заключающей части этой работы производится разработка библиотеки, упрощающая создание такой модели.

Глава 4

Прогнозная способность моделей

При анализе экономических показателей, часто ставится цель прогнозирования будущего состояния временных рядов. На основании этих прогнозов выбираются стратегии дальнейших действий. Исторически, в продвижении этой цели использовались авторегрессионные модели; модели MS-VARX являются расширением этого класса, поэтому возникает вопрос об их прогнозной способности в сравнении с "обычными" методами.

4.1 Задача валидации модели

Для оценивания прогнозной способности моделей линейной регрессии часто используется процедура, которая называется "кросс-валидация" в которой случайная часть тренировочного набора выбирается в качестве тестовой выборки для оценивания поведения модели на "новых" данных. Однако, для моделей временных рядов эта процедура имеет серьезные недостатки. Во-первых, если выбрасывать данные из "середины" временного ряда, то становится невозможным предсказания последующих наблюдений. Во-вторых, из-за серийной корреляции, оценка ошибки предсказания будет значительно ниже чем оно есть на самом деле. Так же возникает проблема с недостатком данных, если есть только одна реализация временного ряда не очень большой длины, и характеристики ряда могут меняться со временем.

Для избегания этих проблем используется процедура "скользящей валидации". Алгоритм состоит в следующем (наглядный пример ниже):

1. Выбор длины валидации L для временного ряда y_t длиной T .
2. Для значений $l \in \{L, L - 1, \dots, 1\}$:
 - (a) Оценивается модель на данных $y_1 \dots y_{T-l}$
 - (b) Вычисляется ошибка прогноза на один шаг вперед: $\hat{\varepsilon}_{T-l+1} = y_{T-l+1} - \hat{y}_{T-l+1}$
3. Считается метрика (часто - среднеквадратичное или среднее абсолютное отклонение) по $\hat{\varepsilon}_{T-L+1} \dots \hat{\varepsilon}_T$, которая и является оценкой этой метрики для будущих прогнозов данного ряда.

Существуют и другие методологии (хороший обзор проведен в [1]), но из-за небольшой длины исходных временных рядов, скользящая валидация считается самой точной.

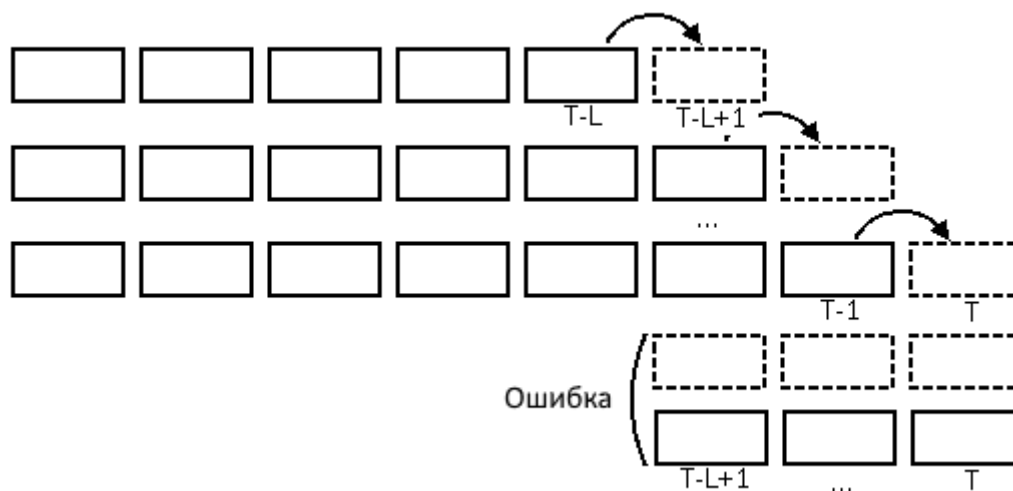


Рисунок 4.1 — Принцип скользящей валидации.

Таблица 4.1 — Сравнение ошибок прогнозов

	Метрика	MS-ARX	SARIMAX
ВВП по Хамильтону	MAE	0.5646	0.0850
	RMSE	0.6019	0.1150
Годовые темпы роста	MAE	0.0988	0.0137
	RMSE	0.1005	0.0164

4.2 Сравнение точности прогнозов моделей для ВВП

Вышеописанный подход "скользящей валидации" был использован для оценки точности прогноза для описанных моделей MS-ARX для годовых темпов роста ВВП (GRGDP) и для ВВП, скорректированного по Хамильтону (HamGDP). В качестве "обычных" методов для обоих рядов рассматривались модели модели *SARIMAX* с автоматически подобранными порядками (алгоритм выбора описан в [10] и имплементировал в пакете Python **pyramid_arima** [16]). В таблице 4.1 указаны среднеабсолютные (MAE) и среднеквадратичные (RMSE) ошибки прогнозов моделей для ВВП, скорректированного по Хамильтону, и для годовых темпов роста ВВП. Ниже представлены графики прогнозов моделей.

Как видно, модели $MS(2) - AR(0)X$ сильно уступают подобранным моделям *SARIMAX*. Можно сделать вывод, что модели переключения среднего не годятся для предсказания ВВП, а только для классификации прошлых периодов (циклов) и оценивания поворотных точек.

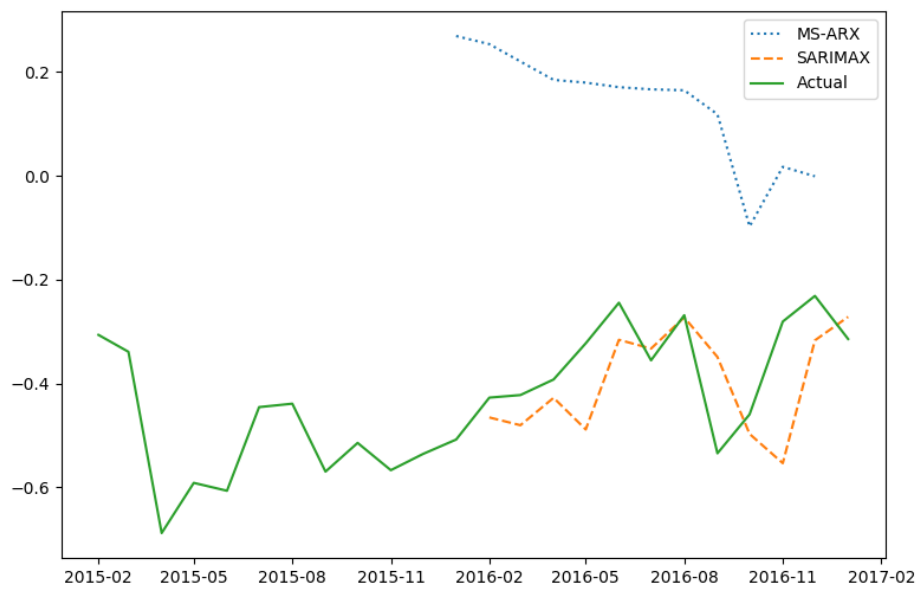


Рисунок 4.2 — Сравнение скользящих прогнозов по Хамильтону.

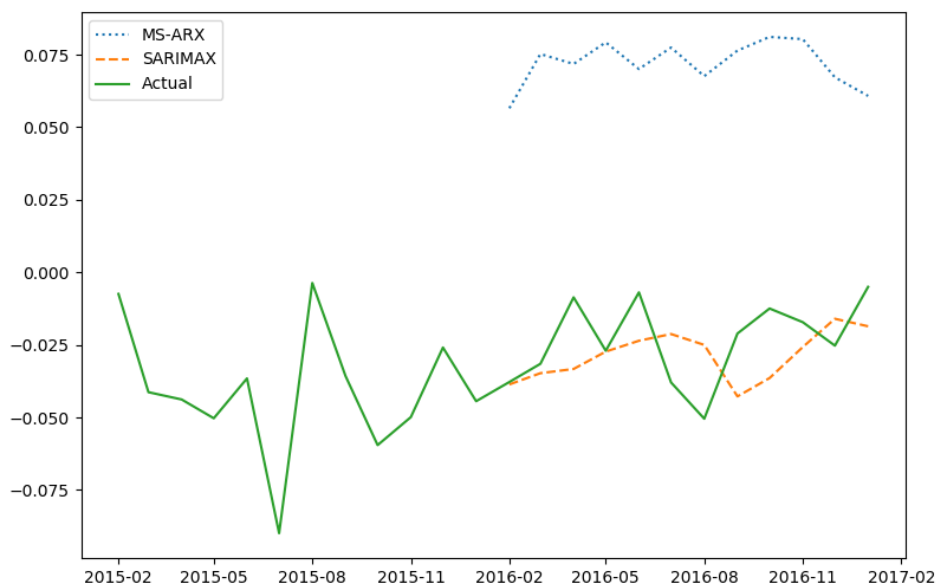


Рисунок 4.3 — Сравнение скользящих прогнозов для темпов роста ВВП.

Глава 5

Разработка библиотеки Time Series for Data Science

Одно из направлений дальнейших исследований – векторные модели переключения состояния (MS-VARX) для отдельных секторов экономики РБ. Для оценивания параметров такой модели уже разработан EM-алгоритм [11], но он не воплощен ни в одном пакете для языков R и Python. Другое направление – оценивание параметров при известных значениях поворотных точек, для чего также не существует пакета.

Реализация алгоритмов оценивания требует больших усилий, включая работы над формализацией входных данных, проверкой предположений и выводом результатов. Для проверки корректности работы этих алгоритмов, необходимо иметь возможность сравнивать их результаты с результатами других моделей существующих пакетов. Наконец, для упрощения сравнения результатов разных моделей удобно представлять их в одинаковом формате.

По этим причинам и на основании опыта работы в поле «data science», автором была разработана библиотека «Time Series for Data Science» (коротко – **ts4ds**) на языке программирования Python.

5.1 Краткое описание и основные задачи

Time Series for Data Science (сокращенно – **ts4ds**, на русском – «Временные Ряды для Анализа Данных») – пакет для языка программирования Python, написанный автором во время практики в компании «ЭПАМ Системз» (EPAM Systems). Он включает в себя средства для анализа и прогнозирования временных рядов, а также средства для разработки новых моделей. С точки зрения программного кода, он содержит базовые классы, разработанные модели, утилиты и программные тесты.

При проектировании и написании пакета **ts4ds**, были поставлены следующие задачи:

- использовать методы статистического и машинного обучения для построения эконометрических моделей временных рядов;
- разработать обобщенное представление модели временных рядов, с целью автоматизации процессов ее построения и применения;
- реализовать в библиотеке процедуры построения, применения некоторых основных семейств моделей

- реализовать процедуры автоматизации работы с моделями (например, поиск гипер-параметров);
- разработать утилиты для облегчения создания новых моделей;
- провести тестирование компонент библиотеки (в программном и в статистическом смысле).

5.2 Принципы разработки библиотеки

Для выполнения поставленных задач был необходим фундаментальный подход к проектированию и разработки библиотеки. В качестве основы брались идеи как из эконометрики, так и из дисциплины машинного обучения. Следует пояснить некоторые термины, взятые из этих дисциплин:

Модель – объект, описывающий (и, возможно, предсказывающий) поведение временных рядов. Математически, модель представляется параметризованными уравнениями, описывающими отношения между переменными. Эти переменные можно условно разбить на **эндогенные** (объясняемые моделью) и **экзогенные** (не объясняемые, а взятые извне и неизменяемые внутри модели).

Предиктивные модели позволяют строить прогнозы значений временных рядов. Примерами являются авторегрессионные модели, экспоненциальное сглаживание, и рекуррентные нейросети.

Преобразования временных рядов изменяют значения временного ряда по определенному принципу, получая при этом новый ряд. Примерами являются взятие разностей, сезонная корректировка, и методы декорреляции. (При разработке библиотеки оказалось, что преобразования удобно рассматривать как вид моделей.)

Параметры модели - обозначения их коэффициентов, а также значения этих коэффициентов. Иногда отдельно выделяют гиперпараметры, которые влияют на саму форму модели, но в пакете **ts4ds** такое отличие не делается.

Эстиматор (от англ. estimator) - алгоритм, который ищет оптимальную (по какому-то критерию) модель. Обычно этот поиск происходит в пространстве параметров, т.е. это алгоритмы оценивания параметров.

Удобно рассматривать множество всевозможных моделей как функциональное пространство зависимостей между переменными. Конкретные классы моделей выделяют подмножества моделей, «точки» в предложенном функциональном пространстве, которые характеризуются определенной параметризацией. Например: множество линейных авторегрессионных моделей скалярной величины $AR(p)$ включают в себя подмножество авторегрессии второго порядка $AR(2)$, которое в свою очередь включает конкретную модель $AR(2)$ с коэффициентами $[-0.61, 0.22]$. В такой

постановке эstimаторы можно рассматривать как алгоритмы нахождения оптимальной точки (комбинации параметров) в некотором подпространстве.

5.3 Особенности имплементации

В библиотеке **ts4ds** каждой из вышеописанных частей соответствует отдельный класс объектов. В рамках пакета данные представляются в виде многомерного массива (`numpy.ndarray`) либо индексированной таблицей (`pandas.Series/DataFrame`) – что является стандартом представления во всех научных библиотеках Python. Все модели – наследники класса `Model` и дополнительных классов (`Predictor`, `Transformer`) с соответствующими параметрами, унаследованные от `Parameters`, и стандартными по интерфейсу процедурами. Эstimаторы (наследованные от `Estimator`) воплощают вышеописанные алгоритмы и возвращают объекты моделей вместе с параметрами.

Стоит отметить, что такая организация несколько отличается от большинства традиционных библиотек. Во многих библиотеках (например, **scikit-learn**, **statsmodels**[15], и многих библиотеках языка R) процедуры оценивания параметров и процедуры выдачи предсказаний совмещены в один объект. Часто в этой структуре заодно хранят и данные. Этот подход обоснован традиционным применением: эконометрическая модель строится на одной реализации процесса и анализируется вручную. Однако, как показывает практика, для многих промышленных применений такой традиционный подход значительно ограничивает потенциал автоматизации.

Главные недостатки традиционной организации (т.к. такой, как в **statsmodels**):

- неудобно сохранять модели компактно, так как нет явного понятия «минимальных данных для воспроизведения»;
- эstimаторы, которые оценивают одинаковые модели (например, МНК и LASSO), должны заново воплотить процедуру предсказания; тем более, могут быть эквивалентные репрезентации, которые усложняют сравнение вычисленных параметров для одинаковых моделей;
- модели «привязаны» к данным, что делает невозможным их применение даже к данным относительно среднего размера, не говоря уже о «больших данных»;
- многие алгоритмы машинного обучения, особенно нейросети, построены с использованием отдельных структур модели, метода оценивания и данных.

Подход привязывания данных легко «включить» в **ts4ds**: для сохранения этой функциональности, а также и краткости записи, существует возможность «привязать» данные к модели через обертку `BoundModel`. Кроме компактной дополнительной нотации в программном коде, особых недостатков у подхода «разделения ролей» нет.

5.4 Результаты разработки

Базовые классы в библиотеке **ts4ds** - `Parameters` (параметры), `Model` (модель), `Estimator` (эстиматор/алгоритм оценивания). Эти классы описывают стандартные интерфейсы для наследующих классов и, совместно с функциями в отделе «devtools», освобождают от рутинной работы (например, проверки входных данных). На их основе воплощены некоторые популярные модели (и эстиматоры для них). Они служат как готовыми продуктами, которые можно использовать в работе аналитика или эконометриста, так и примерами имплементации для разработчиков собственных моделей. Все модели и эстиматоры, а также и часть утилит, автоматически проверяются программными тестами на правильность реализации конвенций интерфейса и на сходство результатов выполнения на тестовых данных.

Объект параметров (`Parameters`) - набор пар ключ-значение, соответствующие именованным параметрам какой-нибудь модели, функции проверки допустимости параметров и другие вспомогательные функции.

Объект модели (`Model`) включает в себе значение параметров и функции предсказаний (`predict`).

Объект эстиматора (`Estimator`) включает в себе настройки/параметры алгоритма оценивания и реализацию этого алгоритма (`fit`), которая принимает набор данных и возвращает объект модели с определенными параметрами.

Преобразования временных рядов наследованы от классов `Model` и `Transformer`. В общем случае, у преобразования существует обратное преобразование, и его форма зависит от преобразуемых данных. Поэтому удобно рассматривать преобразование как вид модели, с соответствующими параметрами (необходимая информация для прямого и обратного преобразования) и эстиматором, который получает эти параметры. Этот подход позволяет реализовывать обратимые преобразования.

В пакете на текущий момент включены следующие преобразования:

- обычные и сезонные разности, для скаляров и векторов, и обратные процедуры интегрирования;
- цепные трансформации, т.е. композиция преобразований.

Планируется добавить следующие преобразования:

- применение скалярных и векторных функций к каждому периоду (и, при возможности, их обратные функции);
- наивная сезонная декомпозиция;
- сезонная корректировка по процедурам X13-ARIMA-SEATS, TRAMO-SEATS;
- процедура выделения тренда по Хамильтону.

Ниже приведен пример работы с преобразованием взятия разности:

```

1 from ts4ds.models.transforms.difference import Diff
2 from ts4ds.models.transforms.chain import Chain_Estimator as Chain
3 from ts4ds.datasets.stata_data import get_air2
4 import matplotlib.pyplot as plt
5 import statsmodels.api as sm
6 plot_acf = sm.graphics.tsa.plot_acf
7
8 y = get_air2()[ 'lnair' ] # aircraft passengers dataset
9
10 # fit and transform
11 dt1, y1 = Diff().fit_transform(y)
12 dt12, y12 = Diff(S=12).fit_transform(y)
13 dt_, y_ = Chain(Diff(), Diff(S=12)).fit_transform(y)
14
15 # plotting
16 fig, axs = plt.subplots(4, sharex=True)
17 axs[0].plot(y)
18 axs[1].plot(y1)
19 axs[2].plot(y12)
20 axs[3].plot(y_)
21 plt.show()
22
23 # Plot final ACF
24 plot_acf(y_[13:], title="Final autocorrelation")
25 plt.show()

```

В пакете **ts4ds** реализованы следующие предсказательные модели и эstimаторы для них:

- линейная регрессия;
- общий вид авторегрессионных моделей;
- модели линейной авторегрессии ARX, VARX;
- общая модель дискретного пространства состояний (state-space models);
- модели ARIMAX, SARIMAX.
- модели с независимыми и Марковскими переключениями состояний (семейства IS-VARX, MS-VARX).

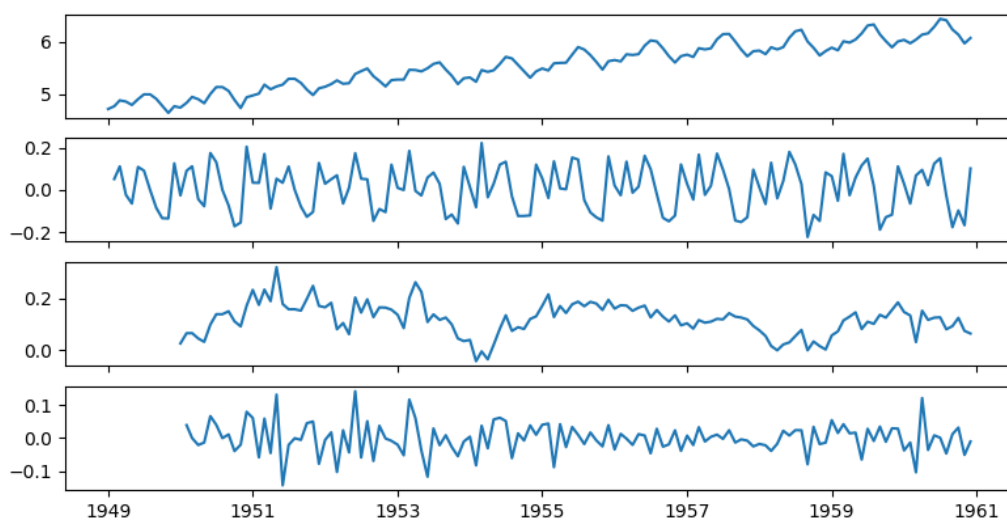


Рисунок 5.1 — График результата выполнения (после различных трансформаций).

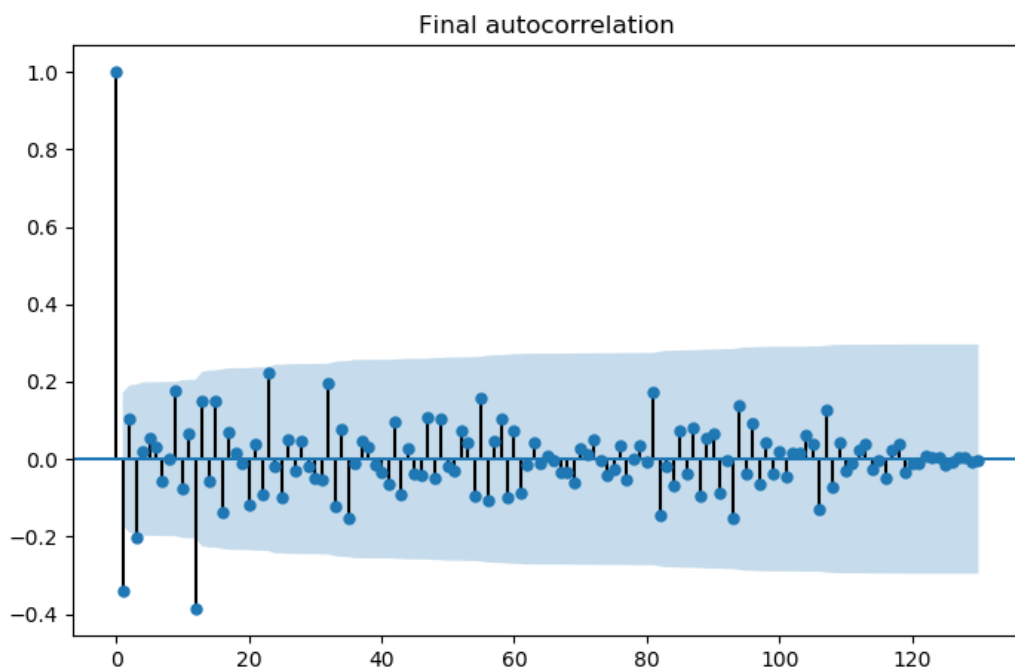


Рисунок 5.2 — График результата выполнения (АКФ после сезонных и первых разностей).

Так же реализованы две процедуры автоматического подбора порядков для SARIMAX, так называемое «auto_arima»[16]. Пример реализации и использования модели SARIMAX описан ниже.

Планируется добавить следующие модели и эstimаторы:

- модели переключения состояний (RS-models), включая MS-VARX;
- модели ARCH, GARCH;
- адаптеры для рекуррентных и сверточных нейросетей;
- адаптер для алгоритмов машинного обучения, изначально предназначенные для пространственных данных (например, SVM и регрессионные деревья).

Пример: разработка модели SARIMAX

Одна из самых распространенных моделей в анализе временных рядов – модель *SARIMAX*, которая является обобщением модели *ARMA*. Полное обозначение этой модели – *SARIMAX*(p, d, q)(P, D, Q, S), где S – порядок сезонности, p и P – порядки обычной и сезонной авторегрессии (*AR*), d и D – порядки обычной и сезонной разности (*I*), q и Q – порядки обычного и сезонного скользящего среднего (*MA*).

Введем обозначения: y_t – эндогенная переменная (в момент времени t), x_t – экзогенный вектор, ε_t – вектор ошибок, L – лаговый оператор, Δ и Δ_s – операторы взятия обычных и сезонных разностей, β – вектор коэффициентов регрессии, $\phi(\cdot)$, $\Phi(\cdot)$, $\theta(\cdot)$, $\Theta(\cdot)$ – многочлены с определенными коэффициентами *AR* и *MA* частей соответственно.

Тогда *SARIMAX*(p, d, q)(P, D, Q, S) можно описать в форме регрессии с остатками *SARIMA*:

$$\begin{cases} y_t = \beta^T x_t + u_t \\ \Phi(L)\phi(L)\Delta_s^Q\Delta^q u_t = \Theta(L)\theta(L)\varepsilon_t \\ \varepsilon_t \sim N(0, \sigma^2) \end{cases}$$

В случае отсутствия экзогенных переменных, $y_t = u_t$.

Для получения предсказаний в такой формулировке необходимо взять d обычных и D сезонных разностей, что эффективно сокращает исходный ряд на $d + SD$ периодов и усложняет задачу предсказания на эти периоды. Вместо этого, можно воспользоваться представлением в форме «пространства состояний» (state-space model) и предсказывать с помощью фильтра Калмана.

Модель state-space описывается данными уравнениями:

$$\begin{cases} \alpha_{t+1} = c + T\alpha_t + R\eta_t \\ y_t = d + Z\alpha_t + \varepsilon_t \\ \eta_t \sim N(0, Q) \\ \varepsilon_t \sim N(0, H) \end{cases}$$

Где y_t - наблюдаемый вектор, α_t - латентный (скрытый) вектор состояний, η_t и ε_t - вектора ошибок в латентном и наблюдаемом векторе соответственно, T , R и Z - матричные параметры модели, c и d - вектора констант, Q и H - ковариационные матрицы.

Представляя в таком виде *SARIMAX*, возможно делать предсказания с помощью фильтра Калмана. Этот подход используется в пакетах **statsmodels**[6] и **ts4ds**. Реализация фильтра Калмана в **statsmodels** очень эффективная, поэтому он используется в качестве основы для предсказаний в **ts4ds**.

Ниже показан пример использования модели SARIMAX для прогнозирования на 4 периода вперед для сезонной модели, а также ряд ошибок предсказаний.

```
1 from ts4ds.estimators.sm_wrap.sarimax import SARIMAX_Estimator
2 from ts4ds.datasets.stata_data import get_air2
3 import matplotlib.pyplot as plt
4 import pandas as pd
5
6 y = get_air2()['lnair'] # airplane passenger dataset
7 oos = pd.date_range('1961-01-01', periods=4, freq='MS') # out-of-sample
8
9 # fit and predict
10 model = SARIMAX_Estimator(p=1, d=1, D=1, S=12).fit(y)
11 y_ins = model.predict_in_sample(y)
12 y_next = model.predict(oos, y)
13 y_hat = pd.concat([y_ins, y_next])
14
15 # plot
16 fig, axs = plt.subplots(2, sharex=True)
17 h1, = axs[0].plot(y, color='0.0')
18 h2, = axs[0].plot(y_hat[13:], color='red', linestyle='—')
19 h4, = axs[1].plot((y_hat-y)[13:], color='0.0')
20 axs[0].legend(handles=[h1, h2, h3], labels=['actual', 'predicted'])
21 plt.show()
```

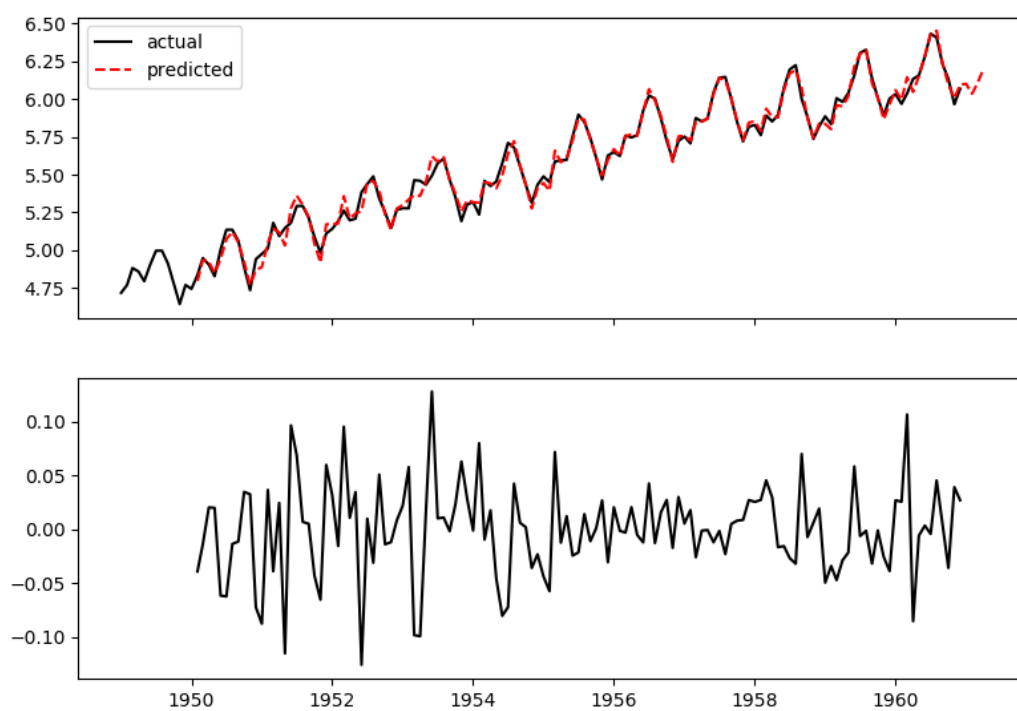


Рисунок 5.3 — Пример работы модели SARIMAX.

Заключение

В работе получены следующие основные результаты:

- исследованы возможности применения методов Ходрика - Прескотта и Хамильтона для выделения циклов в экономических временных рядах;
- на основе этих методов оценены поворотные точки бизнес-цикла белорусской экономики и проведено их сравнение с экспертными оценками;
- построены модели с переключением состояний MS-ARX для временного ряда реального ВВП Республики Беларусь, модели использованы для анализа бизнес-цикла;
- исследованы предиктивные возможности моделей MS-ARX и SARIMAX для реального ВВП;
- спроектирована и разработана библиотека программ **ts4ds** на языке Python для анализа временных рядов ;
- исходный код и данные первых трех глав работы выложены в открытом доступе на сайте <https://github.com/NowanIfideme/PyEconModelling>, исходный код библиотеки на данный момент в закрытом доступе.

Список литературы

1. *Bergmeir C., Hyndman R., Koo B.* A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction. — 2017. — URL: <https://robjhyndman.com/papers/cv-wp.pdf>.
2. *Bodman P., Crosby M.* Phases of the Canadian business cycle // Canadian J. of Economics. — 2000. — Т. 33, № 3. — С. 618—633.
3. *Bruno G., Otranto E.* Models to date the business cycle: The Italian case // Economic Modelling. — 2008. — Т. 25, № 5. — С. 899—911.
4. *Bry G., Boschan C.* Cyclical analysis of time series // National Bureau of Economic Research, Columbia University Press. —
5. *Estrella A.* Extracting Business Cycle Fluctuations: What Do Time Series Filters Really Do. — 2007. — URL: https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr289.pdf (дата обр. 16.05.2016).
6. *Fulton C.* State space modeling in Python. — 2016. — URL: http://www.chadfulton.com/topics/state_space_python.html.
7. *Hamilton J.* A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle // Econometrica. — 1989. — Т. 57, № 2. — С. 357—384.
8. *Hamilton J.* Why You Should Never Use the Hodrick-Prescott Filter. — 01.2017. — Online at Department of Economics, UC San Diego. — Working paper.
9. *Harding D., Pagan A.* A comparison of two business cycle dating methods // J. of Economic Dynamic and Control. — 2003. — Т. 27, № 9. — С. 1681—1690.
10. *Hyndman R., Khandakar Y.* Automatic Time Series Forecasting: The forecast Package for R // Journal of Statistical Software. — 2008. — Т. 27, № 3.
11. *Malugin V., Novopoltsev A.* Statistical Estimation and Classification Algorithms for Regime-Switching VAR Model with Exogenous Variables // Austrian Journal of Statistics. — 2017. — Т. 46. — С. 47—56.
12. *Moore G., Zarnowitz V.* The Development and Role of the National Bureau of Economic Research's Business Cycle Chronologies // The American Business Cycle: Continuity and Change; National Bureau of Economic Research. — 1986. — С. 735—780.
13. *Nilsson R., Gyomai G.* Cycle Extraction. A comparison of the Phase-Average Trend method, the Hodrick-Prescott and Christiano-Fitzgerald filters. — 2011. — URL: <http://dx.doi.org/10.1787/5kg9srt7f8g0-en> (дата обр. 27.05.2016).

14. *OECD*. OECD System of Composite Leading Indicators. — 2012. — URL: <http://www.oecd.org/std/leading-indicators/41629509.pdf> (дата обр. 26.05.2016).
15. *Perktold J., Seabold S., statsmodels-developers*. Python package Statsmodels on statsmodels.org. — 2017. — URL: <http://www.statsmodels.org/stable/index.html>.
16. *Smith T.* Python package pyramid-arma on Github/PyPi. — 2017. — URL: <https://github.com/tgsmith61591/pyramid>.
17. *Statistical Research C. for, Methodology*. X-13ARIMA-SEATS Reference Manual / U.S. Census Bureau. — 2016. — 284 p.
18. *Бабахун Е.* Анализ циклических изменений в экономике Республики Беларусь на основе моделей VAR с Марковскими переключениями состояний : Курсовая работа / Бабахун Е. — БГУ, 2015.
19. *Макаревич А.* Сравнительный анализ оценок поворотных точек экономического цикла на основе алгоритмов Ходрика – Прескотта и Хамильтона // 74-й научная конференция студентов и аспирантов Белорусского государственного университета. — Белорусский Государственный Университет. 2017. — часть 1, с. 53-57.
20. *Макаревич А., Малюгин В.* Построение индекса экономических настроений для Республики Беларусь // Статистические методы анализа экономики и общества: 9-я междунар.науч.-практ. конф. студентов и аспирантов. — Национальный исследовательский университет «Высшая школа экономики». 2018. — Статистические методы анализа экономики и общества: 9-я междунар.науч.-практ. конф. студентов и аспирантов (в печати).
21. *Малюгин В.* Методы анализа многомерных эконометрических моделей с неоднородной структурой. Научное издание. — 2014. — 351 с.
22. Модельные и инструментальные средства для построения и применения индекса экономических настроений белорусской экономики / В. Малюгин [и др.] // Проблемы прогнозирования и государственного регулирования социально-экономического развития: материалы XVII Международной конференции. Т. 1. — НИЭИ Минэкономики Республики Беларусь. 2017. — С. 178—188.
23. Разработка системы опережающих экономических индикаторов и экономических диффузных индексов для основных видов экономической деятельности и экономики Республики Беларусь в целом с использованием экономико-математических, эконометрических методов и моделей на основе данных системы мониторинга предприятий Национального банка Республики Беларусь: отчет о НИР (заключ.) /

НИИ ППМИ; рук. В.И. Малюгин. — Минск, 2017. — 142 с. — ГР
20162817.