

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Белорусский государственный университет

Республиканский конкурс научных работ студентов

Секция 10. Математика. Методы и алгоритмы вычислительной математики
математического моделирования для решения задач экономики, техники и
природоведения

**“РАЗРАБОТКА И ПРИМЕНЕНИЕ ЭКОНОМЕТРИЧЕСКИХ
МОДЕЛЕЙ И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ
АНАЛИЗА ЭКОНОМИЧЕСКИХ ЦИКЛОВ”**

Макаревич Анатолий
Сергеевич,
выпускник кафедры
математического
моделирования и анализа
данных (ММАД)

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

Минск 2018

Реферат

Конкурсная работа: 45 с., 18 рис., 3 табл., 35 источников (5 авторских)

Ключевые слова: ЭКОНОМИЧЕСКИЕ ЦИКЛЫ, ПОВОРОТНЫЕ ТОЧКИ, ИНДЕКС ЭКОНОМИЧЕСКИХ НАСТРОЕНИЙ, ФИЛЬТР ХОДРИКА – ПРЕСКОТТА, ФИЛЬТР ХАМИЛЬТОНА, ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ, МОДЕЛИ С ПЕРЕКЛЮЧЕНИЕМ СОСТОЯНИЙ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

В работе рассматривается задача оценивания поворотных точек (моментов смены фаз) экономического цикла с помощью двух подходов: статистических фильтров, применяемых для выделения трендовых и циклических компонент временного ряда, и эконометрических моделей с переключением состояний. Исследуется эффективность двух подходов при решении задачи анализа бизнес-цикла белорусской экономики на основе временных рядов реального ВВП и индекса экономических настроений Республики Беларусь, построенного по опросным данным белорусских предприятий. В рамках первого подхода проводится сравнительный анализ двух статистических фильтров: фильтра Ходрика – Прескотта (традиционный подход) и нового фильтра, предложенного Дж. Хамильтоном в 2017 г., не имеющего широкого практического применения. В рамках второго подхода построена модель с марковскими переключениями состояний из семейства MS-VARX, включающая в качестве экзогенной переменной индекс экономических настроений. В работе представляются результаты разработки библиотеки программ на языке Python для анализа временных рядов с элементами машинного обучения, включающего использованные в работе эконометрические модели и методы анализа. Эта библиотека на данный момент используется в компании «ИООО ЭПАМ Системз», и идет процесс предоставления исходного кода в открытый доступ.

Содержание

Введение	5
1 Эконометрические модели и методы анализа циклических изменений в экономике	8
1.1 Использование статистических фильтров для выделения трендов и циклов из экономических временных рядов	8
1.2 Общее описание используемых статистических фильтров	10
1.2.1 Фильтр Ходрика – Прескотта	10
1.2.2 Фильтр Хамильтона	12
1.3 Модели семейства MS-VARX и их применение для анализа циклических изменений	14
1.3.1 Общее описание моделей с переключениями состояний .	14
1.3.2 Модель MS-ARX	15
2 Экспериментальное исследование эконометрических методов анализа экономических циклов	16
2.1 Результаты сравнительного анализа фильтров	16
2.1.1 Выделение тренда и циклической составляющей с шумовым компонентом	16
2.1.2 Выделение циклов и анализ поворотных точек	19
2.1.3 Выводы	22
2.2 Анализ циклических изменений и оценка поворотных точек экономических циклов на основе модели с переключением состояний	23
2.2.1 Результаты построения и применения модели MS-ARX .	23
2.2.2 Сравнительный анализ поворотных точек на основе фильтров и модели MS-ARX	25
2.3 Прогностическая способность моделей	28
2.3.1 Задача валидации модели	28
2.3.2 Сравнение точности прогнозов моделей для ВВП	29
3 Разработка библиотеки Time Series for Data Science	32
3.1 Краткое описание и основные задачи	32
3.2 Принципы разработки библиотеки	33
3.3 Особенности имплементации	34
3.4 Результаты разработки	35
3.5 Пример: разработка модели SARIMAX	38
Заключение	41

Список источников	42
Список публикаций автора	45
Приложения. Научные публикации автора.	46

Обозначения и сокращения

ВВП / GDP – валовой внутренний продукт (Gross Domestic Product);
ИЭН / ESI – индекс экономических настроений (Economic Sentiment Index);

ОЭСР / OECD – Организация экономического сотрудничества и развития (Organization for Economic Cooperation and Development);

НБЭИ / NBER – Национальное бюро экономических исследований США (National Bureau of Economic Research)

X13-ARIMA-SEATS и TRAMO/SEATS – методы сезонной корректировки временных рядов;

MS-ARX – модель авторегрессии с Марковским переключением состояний и экзогенными переменными (Markov-Switching AutoRegression with Exogenous variables);

SARIMAX – сезонная, интегрированная модель авторегрессии и скользящего среднего с экзогенными переменными (Seasonal AutoRegressive Integrated Moving Average model with Exogenous variables).

Введение

Анализ экономических циклов и оценка моментов смены фаз циклов, называемых поворотными точками, является одной из актуальных проблем макроэкономического прогнозирования. Для проведения такого анализа используется некоторый базовый экономический индикатор, характеризующий состояние экономики, относительно которого рассчитываются показатели ее роста и спада. В качестве такого индикатора обычно используется реальный ВВП [16; 17; 30].

Можно выделить три главных аспекта проблемы анализа экономических циклов (бизнес-циклов): экономическая интерпретация и обоснование экономических циклов, которые даются в рамках экономической теории; статистические методы выделения долгосрочных циклов их макроэкономических временных рядов; методы оценивания поворотных точек на основе их статистического и экономического обоснования.

Теория экономических циклов

В рамках концепции экономического цикла, используемой в НБЭИ (Национальное бюро экономических исследований) США, подразумевается последовательная смена двух фаз базового экономического индикатора, 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 [15]. 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) [16]. 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) [17]. В рамках [27–29] была разработана система опережающих индикаторов для Республики Беларусь, в том числе Индекс Экономических Настроений (Economic Sentiment Indicator – ИЭН/ESI). В настоящей работе в качестве опережающего и базового экономических индикаторов используются соответственно временные ряды значений индекса экономических настроений и реального ВВП Республики Беларусь в ценах 2014 г., в месячном

исчислении с мая 2005 г. по январь 2017 г., полученные в рамках указанной НИР.

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 реального ВВП. В настоящее время, основным методом выделения догосрочной циклической составляющей являются статистические фильтры. Наиболее популярным является фильтр Ходрика – Прескотта [5; 16]. В недавней статье Дж. Хамильтона [8] указывается на проблемы, возникающие при использовании фильтра Ходрика – Прескотта, и предлагается свой метод выделения долгосрочного цикла. Этот метод, так называемый «фильтр Хамильтона», также подвергается некоторой критике.

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 популяризированные работой Хамильтона [7]. Результаты Хамильтона с MS-AR для бизнес-цикла США во многом сходились с периодизацией NBER.

Содержание данной работы

В рамках данной работы для оценивания и прогнозирования моментов смены фаз реального ВВП применяются два подхода. Первый подход основан на построении и применении опережающих экономических индикаторов. Индикатор считается опережающим, если поворотные точки его цикла (точки переключения фаз цикла «спад» и «подъем») опережают поворотные точки цикла базового экономического индикатора.

В рамках данного подхода ключевыми являются две задачи: задача сезонной корректировки временных рядов базового и опережающего индикаторов, и задача выделения циклической составляющей из сезонно-скорректированных временных рядов с помощью специальных статистических фильтров. На основе сравнения циклов для базового и опережающего индикаторов осуществляется оценка и прогнозирование поворотных точек базового индикатора по поворотным точкам опережающего индикатора.

Второй подход основан на применении эконометрических моделей с марковскими переключениями состояний. Обычно используются модели

одномерной и векторной авторегрессии с марковскими переключениями состояний (MS-VARX) [Ссылки]. Для оценки поворотных точек в рамках данного подхода применяются алгоритмы совместного оценивания номеров классов состояний экономики («спад»/«подъем») и параметров моделей.

В данной работе исследуется эффективность двух подходов при решении задачи анализа бизнес-цикла белорусской экономики на основе временных рядов реального ВВП и индекса экономических настроений Республики Беларусь, построенного по опросным данным белорусских предприятий. В рамках первого подхода проводится сравнительный анализ двух статистических фильтров: фильтра Ходрика – Прескотта (традиционный подход) и нового фильтра, предложенного Дж. Хамильтоном в 2017 г., не имеющего широкого практического применения. В рамках второго подхода строится модель с марковскими переключениями состояний из семейства MS-VARX, включающая в качестве экзогенной переменной индекс экономических настроений.

Работа продолжает исследования, проводившиеся в данном направлении 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"[27]. Решаемые задачи и полученные результаты являются новыми и актуальными для анализа сложных систем с различными классами состояний. В работе также представляются результаты разработки пакета программ на языке Python для анализа временных рядов с элементами машинного обучения, включающего использованные в работе эконометрические модели и методы анализа.

Глава 1

Эконометрические модели и методы анализа циклических изменений в экономике

1.1 Использование статистических фильтров для выделения трендов и циклов из экономических временных рядов

Статистический фильтр Ходрика – Прескотта [12] широко используется в различных эконометрических исследованиях, связанных с выделением долгосрочных трендов и циклов из экономических временных рядов. Выделение трендовой компоненты исследуемого временного ряда является важным этапом при решении такой задачи макроэкономического анализа, как определение потенциального выпуска и разрыва выпуска (кредитного разрыва) в рамках формирования денежно-кредитной политики центральными банками [20; 24; 25]. При этом трендовая компонента временного ряда рассматривается как потенциальный выпуск или равновесный реальный ВВП, а остаточная составляющая временного ряда, включающая циклический и шумовой компонент, соответствует отклонениям его фактических значений от ожидаемых в равновесном состоянии и интерпретируется как разрыв выпуска. В рассматриваемом контексте важной является задача прогнозирования значений потенциального выпуска и разрыва выпуска в экономике на ближайшую перспективу.

Фильтр Ходрика – Прескотта в настоящее время также является рекомендуемым методом выделения долгосрочных циклов в рамках методологии ОЭСР [16; 17], используемой для анализа бизнес-циклов на основе опережающих экономических индикаторов. В контексте подобного исследования он применялся для выделения циклических компонент из временных рядов реального ВВП и индекса экономических настроений (ИЭН) белорусской экономики [28].

Однако, широко используемый стандартный одномерный фильтр Ходрика – Прескотта обладает рядом общеизвестных недостатков. К ним относятся:

- смещение оценок значений тренда в начальных и конечных точках временного ряда («end-point bias problem») и высокая чувствительность этих оценок к добавлению новых наблюдений;
- отсутствие строгой формализации правила выбора для значения параметра фильтра («параметра λ »), определяющего степень гладкости

тренда и критическим образом влияющего на свойства выделяемых компонентов временного ряда.

Отметим, что смещение имеет место и в начальных точках временного ряда, но оно не является критичным в контексте задачи прогнозирования.

Также имеют место особенности применения данного фильтра к временным рядам, в зависимости от того, содержат ли они детерминированные или стохастические тренды [30]. Это можно объяснить тем, что в основе процедуры выделения тренда лежит компромисс между степенью гладкости получаемого детерминированного тренда и его близости к фактическим данным. При этом ожидается, что остаточная составляющая представляет собой комбинацию цикла и шумового компонента, который является стационарным временным рядом и имеет нулевое среднее значение. Можно предположить, что указанные свойства трендовой и остаточной составляющих достижимы, если сглаживаемый временной ряд относится к классу нестационарных временных рядов с детерминированными трендами (trend stationary – TS-моделей). В противном случае, если временной ряд содержит стохастический тренд, т. е. относится к классу DS-моделей (difference stationary), следует ожидать нестационарность шумового компонента в остаточной составляющей, следствием которой являются ложные циклы (spurious cycles) и ложные корреляции (spurious correlations). Эти недостатки отмечаются исследователями при обработке большей части макроэкономических и финансовых временных рядов [10; 18; 20], проявляющих свойства процессов случайного блуждания.

В этом контексте Дж. Хамильтон выступил с критикой использования фильтра Ходрика – Прескотта в качестве универсального подхода к выделению долгосрочных трендов. В своей рабочей статье он приводит аналитическое обоснование следующих недостатков фильтра [8]:

1. фильтр индуцирует ложные циклы, когда применяется к временным рядам с высоким порядком интегрирования и содержащим стохастический тренд;
2. конечные значения тренда, выделяемого из временного ряда, существенно отличаются от срединных значений и описывают ложную динамику;
3. статистическая формализация проблемы выбора параметра сглаживания тренда обычно приводит к значениям параметра, который в значительной степени противоречит обычной практике.

В этой же статье автором предлагается альтернативный метод выделения тренда, получивший в литературе название «регрессионный фильтр Хамильтона» (Hamilton's regression filter). К настоящему моменту

известен ряд публикаций, посвященных сравнительному анализу фильтров Ходрика – Прескотта и Хамильтона в практических исследованиях. В рабочей статье [20] проведен сравнительный анализ обоих фильтров при анализе «кредитного разрыва» в экономике (credit-to-GDP gap) на основе квартальных временных рядов США. В работе [11] при анализе цикличности банковского капитала использовались оба фильтра и в целом получены схожие результаты.

Фильтр Хамильтона, в отличие от фильтра Ходрика – Прескотта, еще не использовался белорусскими исследователями. Поэтому представляется актуальным сравнительный анализ обоих фильтров при обработке белорусских макроэкономических временных рядов, обладающих такими особенностями, как короткая длина и частые структурные изменения. Данная статья посвящена сравнительному анализу результатов применения фильтров для выделения трендовых и циклических компонентов временных рядов при решении задачи анализа бизнес-цикла белорусской экономики в рамках методологии ОЭСР [16; 27].

1.2 Общее описание используемых статистических фильтров

1.2.1 Фильтр Ходрика – Прескотта

При использовании данного фильтра априорно предполагается, что временной ряд имеет структуру, допускающую его нестационарность в виде наличия тренда, а также присутствие циклических изменений. Таким образом, общая структура временного ряда допускает представление:

$$y_t = \tau_t + c_t + \varepsilon_t, \quad t \in \overline{1, T}, \quad (1.1)$$

где временные ряды τ_t , c_t и ε_t соответствуют трендовой, циклической и случайной шумовой компонентам.

Обычно выделению тренда предшествует сезонная корректировка временного ряда. Для этой цели могут применяться известные процедуры Census X13, TRAMO-SEATS и др. [16; 28].

Фильтр Ходрика – Прескотта, при однократном применении, осуществляет декомпозицию временного ряда y_t на тренд τ_t и остаточную составляющую ν_t вида

$$y_t = \tau_t + \nu_t. \quad (1.2)$$

Остаточная составляющая включает цикл, искаженный случайным шумом, т. е.

$$\nu_t = c_t + \varepsilon_t. \quad (1.3)$$

Для выделения тренда τ_t решается задача оптимизации [12]:

$$\min_{\tau_t} \left(\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} (\tau_{t+1} - 2\tau_t + \tau_{t-1})^2 \right), \quad \lambda > 0, \quad (1.4)$$

где $\lambda > 0$ – параметр сглаживания тренда: при $\lambda \rightarrow 0$ значения тренда близки к значениям исходного ряда, т. е. $\tau_t \rightarrow y_t$, а при $\lambda \rightarrow \infty$ вид тренда приближается к линейной по времени t функции.

Первое слагаемое в уравнении (1.4) отвечает за точность подгонки, а второе – за степень гладкости тренда. Как отмечалось выше, возникает проблема выбора значения параметра для конкретных условий решаемой задачи. На практике рекомендуются следующие статистически обоснованные значения параметра в зависимости от интервала наблюдения значений временного ряда: 14400 – для месячных, 1600 – для квартальных и 100 – для годовых временных рядов [12]. Однако, на практике значения корректируются с учетом особенностей решаемой задачи, длины предполагаемых циклов и используемых временных рядов.

Так, например, при анализе «кредитного разрыва» в экономике (credit-to-GDP gap) на основе квартальных временных рядов достаточно большой длины в соответствии с рекомендациями Базель-3 следует полагать $\lambda = 400000$ что соответствует существованию циклов длиной 30 лет. В то время как при $\lambda = 1600$ ожидаемая длина цикла составляет около 7.5 лет [20].

Фильтр Ходрика – Прескотта обладает «свойством симметричности», которое проявляется в использовании начальных и конечных значений временного ряда в процессе сглаживания его срединных значений; иными словами, он является двусторонним фильтром. Следствием данного свойства является смещенность начальных и конечных сглаженных значений. В контексте задачи прогнозирования особую важность приобретает проблема смещения конечных значений тренда. Данная проблема в стандартной версии фильтра решается путем предварительной пролонгации временного ряда (например, с помощью прогнозирования на основе модели ARIMA). Таким образом, интересующие пользователя точки будут находиться уже не в конце временного ряда, а искажению вместо них подвергнутся только дополнительно пролонгированные значения. Удобным способом продления временного ряда длиной T на заданную величину H с его одновременной сезонной корректировкой является применение метода TRAMO/SEATS для расширенного временного интервала длиной $T+H$. Такая возможность имеется во многих статистических и эконометрических пакетах, в том числе в программе ESI Analysis (Economic Sentiment Indicator Analysis), разработанной на языке программирования R в рамках научного проекта,

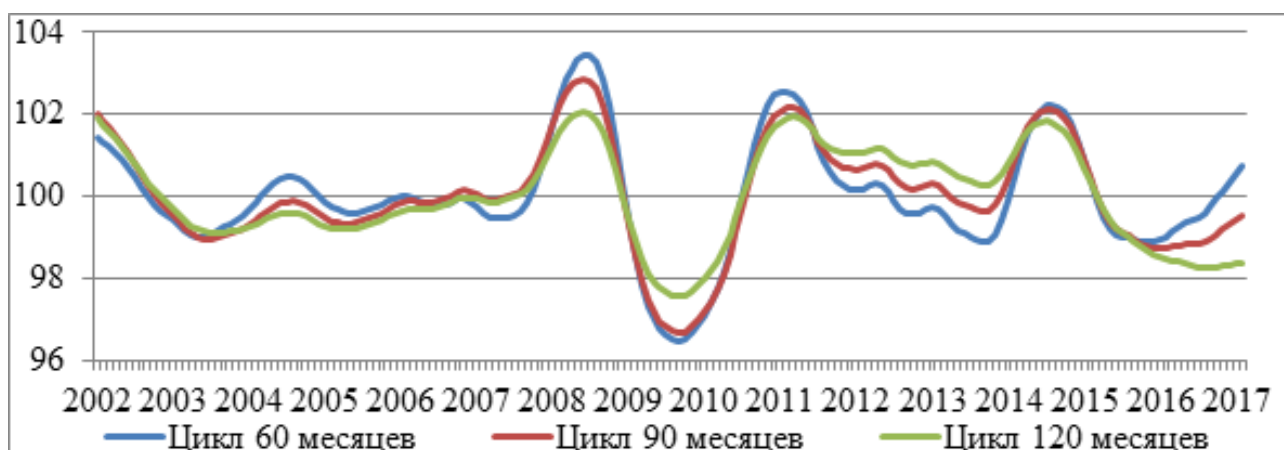


Рисунок 1.1 — Оценки бизнес-цикла посредством фильтра Ходрика – Прескотта в зависимости от предполагаемой продолжительности цикла.

выполнявшегося в НИИ ППМИ БГУ по заданию Национального банка Республики Беларусь в 2016-2017 гг. [27]. Программа ESI Analysis реализует весь комплекс задач, связанных с построением индекса экономических настроений и его применения для анализа бизнес-цикла белорусской экономики и его поворотных точек.

Для выделения долгосрочного цикла из временного ряда y_t используется процедура, основанная на двухэтапном применении фильтра Ходрика – Прескотта. Для этой цели к остаточной составляющей ν_t , полученной после выделения тренда из временного ряда y_t , применяется фильтр Ходрика – Прескотта со значением λ , соответствующим более краткосрочным колебаниям. В рамках данного исследования, при анализе бизнес-цикла в белорусской экономике на основе месячных временных рядов, используются рекомендуемые в рамках методики ОЭСР значения параметра λ : $\lambda = 42131.155$ для первого и $\lambda = 13.93$ для второго этапа процедуры. Этим значениям соответствуют продолжительности «типичного бизнес-цикла» (от 5 до 8-9 лет) и высокочастотных колебаний (6-12 месяцев). На рисунке 1.1 представлены оценки циклической составляющей белорусского ВВП с различными величинами параметра λ (соответствуют длине цикла 60, 90 и 120 месяцев) [28]. Отличия в этих оценках циклических составляющих невелики, за исключением конечных значений. В описываемых ниже экспериментах используется средний по продолжительности цикла вариант, которому соответствуют вышеуказанные значения параметра λ .

1.2.2 Фильтр Хамильтона

В фильтре, предложенном Дж. Хамильтоном [8], долгосрочный тренд в нестационарном временном ряду y_t описывается моделью регрессии, включающей, для момента времени $t + h$, константу и четыре наиболее

близких к моменту времени t значения временного ряда:

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + \nu_{t+h}, \quad (1.5)$$

где ($h > 0$) – априорно задаваемый параметр, зависящий от предполагаемой длины цикла.

При анализе бизнес-цикла по квартальным данным США в [8] предлагается использовать, $h = 8$ (для двухгодичных циклов), а в исследованиях кредитных и финансовых циклов $h = 20$ (для пятилетних циклов).

Для оценивания параметров модели (1.5) используется обычный метод наименьших квадратов (МНК). Остаточная составляющая, появляющаяся после подстановки в регрессионное уравнение (1.5) МНК-оценок параметров $\{\hat{\beta}_l\}$, $l = \overline{0, 4}$, описывается соотношением:

$$\hat{\nu}_{t+h} = y_{t+h} - \hat{\beta}_0 - \hat{\beta}_1 y_t - \hat{\beta}_2 y_{t-1} - \hat{\beta}_3 y_{t-2} - \hat{\beta}_4 y_{t-3}. \quad (1.6)$$

Хамильтоном показано, что для нестационарных интегрированных временных рядов вплоть до порядка $d = 4$ остаточная циклическая составляющая ν_t временного ряда вида (1.6) включает как циклическую, так и стационарную случайную компоненту, т. е. $\hat{\nu}_{t+h} = c_{t+h} + \varepsilon_{t+h}$. При тех же условиях остаточная составляющая в уравнении (1.2), получаемая с помощью фильтра Ходрика – Прескотта, может демонстрировать ложные циклы и корреляции.

Достоинством процедуры Хамильтона является то, что она не использует априорную информацию о конкретном виде модели временного ряда, например о порядке интегрируемости.

В частном случае, если временной ряд порождается моделью случайного блуждания, т. е. порядок интегрирования ряда равен 1:

$$y_t = \beta_0 + y_{t-1} + \varepsilon_t, \quad (1.7)$$

то для МНК-оценок коэффициентов регрессии в уравнении (1.6) при достаточно большой длине ряда ($T \rightarrow \infty$) значения сходятся к $\beta_1 = 1$ и $\beta_l = 0, l \in \{2, 3, 4\}$ [8, с. 16-17]. В данном случае исключение тренда по Хамильтону эквивалентно взятию первой разности (со смещением):

$$\hat{\nu}_{t+h} = y_{t+h} - \hat{\beta}_0 + y_t. \quad (1.8)$$

Для временного ряда с порядком интегрирования 2 при оценке модели (1.5) следует ожидать получение статистически значимого значения коэффициента регрессии $\hat{\beta}_3$ (для лага -2).

Указанными свойствами обладают так называемые *разностные фильтры* (difference filters), т. е. фильтры, основанные на взятии разностей; следовательно, результаты применения фильтра Хамильтона к временным

рядам с невысоким порядком интегрирования могут быть близки к результатам, получаемым с помощью разностных фильтров. В отношении разностных фильтров стоит отметить, что их не следует применять к стационарным временным рядам или временным рядам, содержащим детерминированные тренды из-за возможной генерации ложных циклов [10; 20]. Для подобного же утверждения в отношении свойств фильтра Хамильтона для таких рядов требуются дальнейшие исследования.

В отличие от фильтра Ходрика – Прескотта, фильтр Хамильтона является «асимметричным» (т. е. односторонним), что позволяет избежать проблемы смещения конечных значений при прогнозировании остаточной составляющей. Так, например, для временного ряда, описываемого моделью случайного блуждания (1.7), значение остаточной циклической составляющей $\hat{\nu}_{t+h}$ в момент времени $t+h$ формируется на основе известных значений временного ряда y_{t+h} и y_t . В общем случае для оценивания тренда и получения остаточной циклической составляющей используется только наблюдаемый временной ряд; это становится особенно важным для коротких временных рядов, а также при наличии структурных изменений в анализируемом временном интервале. Более глубокий сравнительный анализ свойств обоих фильтров проводится в [20].

Далее рассматривается альтернативный метод оценки поворотных точек на основе моделей с переключением состояния.

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

1.3.1 Общее описание моделей с переключениями состояний

Модели с переключением состояний (Regime Switching / *RS* models) – это подкласс моделей со структурными изменениями в которых скачкообразно меняются параметры модели. Каждый «режим» (класс состояния) в данном случае соответствует отдельному набору параметров, и номер режима $l(t)$ в текущий момент времени t является ненаблюдаемой случайной величиной.

Если номер режима $l = l(t)$ описывается Марковским процессом с L классов состояний (текущее состояние зависит только от предыдущего, с матрицей вероятностей перехода $L \times L$), то такую модель называют *марковской моделью переключения состояний* (Markov Switching / *MS* model). Подробнее эти модели описаны в [14; 26].

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

$$y_t = \alpha_{0,l} + \sum_{i=1}^p A_{i,l} y_{t-i} + B_l x_t + \eta_t, \quad t = \overline{1, T} \quad (1.9)$$

где p – порядок авторегрессии, x_t – экзогенная векторная переменная, $\alpha_{0,l}$ – вектор констант регрессии, 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 одношаговых переходов для цепи Маркова:

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,L} \\ m_{2,1} & m_{2,2} & \dots & m_{2,L} \\ \dots & \dots & \dots & \dots \\ m_{L,1} & m_{L,2} & \dots & m_{L,L} \end{bmatrix}, \quad \sum_{j=1}^L m_{i,j} = 1 \quad \forall i \in \overline{1, L}. \quad (1.10)$$

При $L = 2$, можно упростить матрицу:

$$M = \begin{bmatrix} \sigma_1 & 1 - \sigma_2 \\ 1 - \sigma_1 & \sigma_2 \end{bmatrix}. \quad (1.11)$$

Оценивание параметров таких моделей можно проводить с помощью итерационного ЕМ–алгоритма (ЕМ MS–VARX) [14] либо методом максимального правдоподобия. Из–за стохастичного характера алгоритма сходимость к одному результату не гарантирована.

1.3.2 Модель MS-ARX

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

$$y_t = \alpha_{0,l} + \sum_{i=1}^p \alpha_{i,l} y_{t-i} + \beta_l x_t + \eta_t, \quad (1.12)$$

где η_t – одномерный гауссовский белый шум с матрицей σ_l . Параметры σ_l , $\alpha_{0,l}$, $\alpha_{i,l}$ и β_l зависят от режима l .

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

$$y_t = \alpha_{0,l} + \sum_{i=1}^p \alpha_{i,l} y_{t-i} + \beta_l x_{t-k} + \eta_t. \quad (1.13)$$

Обозначим эту модель $MS(L) - AR(p) - X(x_k)$. Оценивание можно проводить по вышеописанными алгоритмами для MS-VARX.

Глава 2

Экспериментальное исследование эконометрических методов анализа экономических циклов

2.1 Результаты сравнительного анализа фильтров

2.1.1 Выделение тренда и циклической составляющей с шумовым компонентом

Приведем описание результатов применения фильтров Ходрика – Прескотта и Хамильтона для решения задачи анализа бизнес-цикла белорусской экономики в рамках методологии ОЭСР [16]. Статистической обработке с помощью указанных фильтров подвергаются два временных ряда, полученные за период наблюдения с мая 2005 г. по январь 2017 г.:

1. базовый экономический индикатор, на основании которого определяется бизнес-цикл белорусской экономики, в виде реального ВВП в ценах 2014 г.;
2. индекс экономических настроений ИЭН, построенный на основе данных конъюнктурных опросов предприятий в рамках системы мониторинга Национального банка Республики Беларусь.

Для обработки временных рядов в настоящем исследовании используется пакет `statsmodels` языка Python [19].

Методология построения и результаты сравнительного анализа циклов данных индикаторов описаны в [28]. Там же осуществлена оценка поворотных точек для обоих циклов и установлен опережающий характер индекса ESI с периодом опережения в среднем на 4 месяца. Для выделения трендов и циклов из обоих временных рядов в соответствии с методологией ОЭСР [20] использовался фильтр Ходрика – Прескотта. В связи с появлением нового метода решения данных задач в виде фильтра Хамильтона становится актуальной задача сравнительного анализа результатов, полученных с помощью фильтров Ходрика – Прескотта и Хамильтона. В данном разделе представляются результаты проведенного сравнения.

На рисунке 2.1 представлены используемые месячные временные ряды реального ВВП и ИЭН Республики Беларусь. В соответствии с вышеописанными процедурами обработки временных рядов на

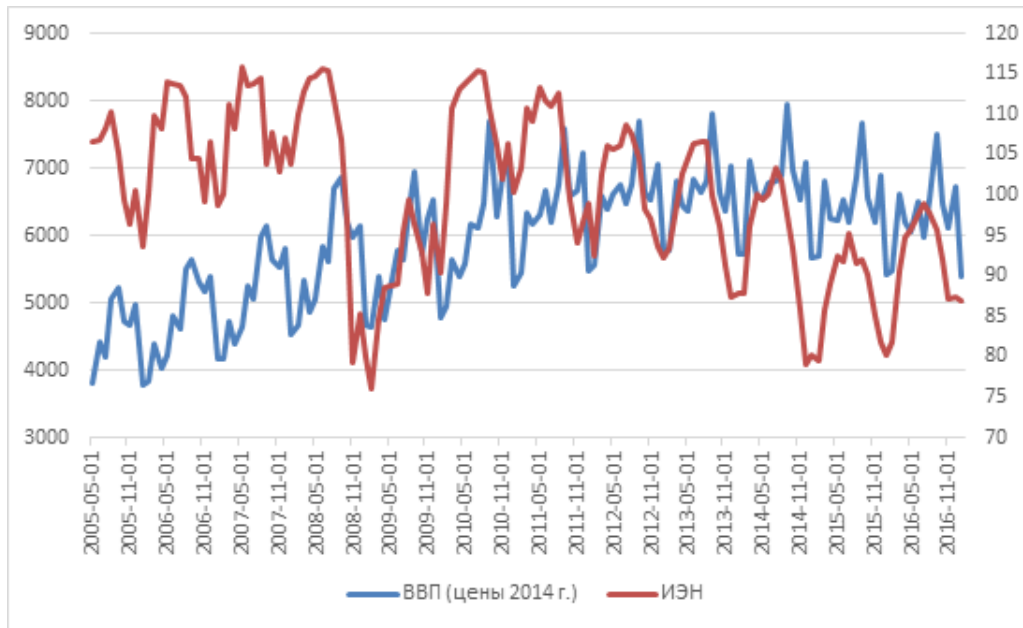


Рисунок 2.1 — Временные ряды ВВП и ИЭН. Шкала для ВВП — слева, для ИЭН — справа.

предварительном этапе осуществлялась сезонная корректировка временных рядов методом X13-ARIMA-SEATS [22].

Далее из сезонно скорректированных временных рядов ВВП и ИЭН выделяются трендовые компоненты с применением каждого из фильтров. Параметры сглаживания трендов полагаются следующими: $\lambda = 42131.155$ для фильтра Ходрика – Прескотта [28] и $h = 12$ для фильтра Хамильтона. Данное значение параметра h , соответствующее годовичному циклу по месячным данным, выбрано в соответствии с общими рекомендациями в рамках фильтра Хамильтона и с учетом относительно короткой длины рассматриваемых временных рядов. Полученные тренды и сезонно скорректированные временные ряды для реального ВВП и ИЭН представлены на рисунках 2.2 и 2.3.

В результате оценивания регрессионных моделей вида (1.5) получены следующие уравнения трендов для реального ВВП (2.1) и ИЭН (2.2). В квадратных скобках указаны значения t-статистики для теста статистической значимости соответствующих оценок параметров модели, $\hat{\sigma}^2$ представляет собой оценку дисперсии остатков:

$$y_{t+12} = 1490.67 + 0.918y_t - 0.017y_{t-1} - 0.106y_{t-2} - 0.018y_{t-3} + \hat{\nu}_{t+h}, \quad \hat{\sigma}^2 = 5851, \quad (2.1)$$

$\begin{matrix} [7.63] & [23.21] & [-0.43] & [-2.68] & [-0.47] \end{matrix}$

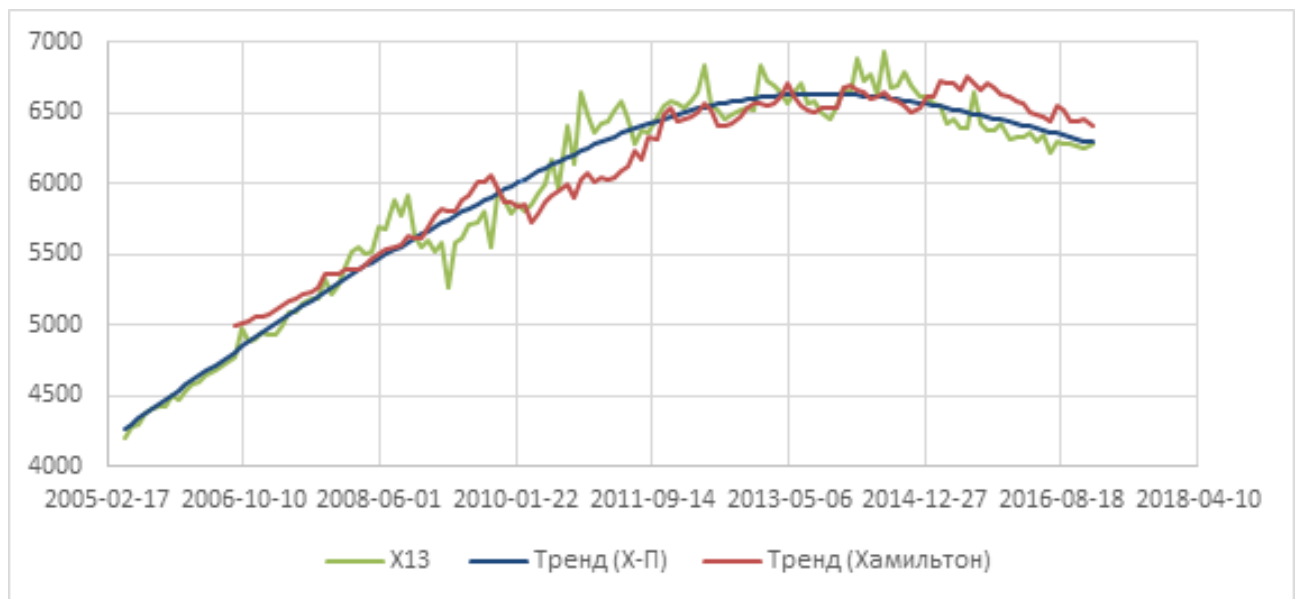


Рисунок 2.2 — Сезонно скорректированный ряд ВВП и тренды, выделенные фильтрами Ходрика – Прескотта (Х-П) и Хамильтона.

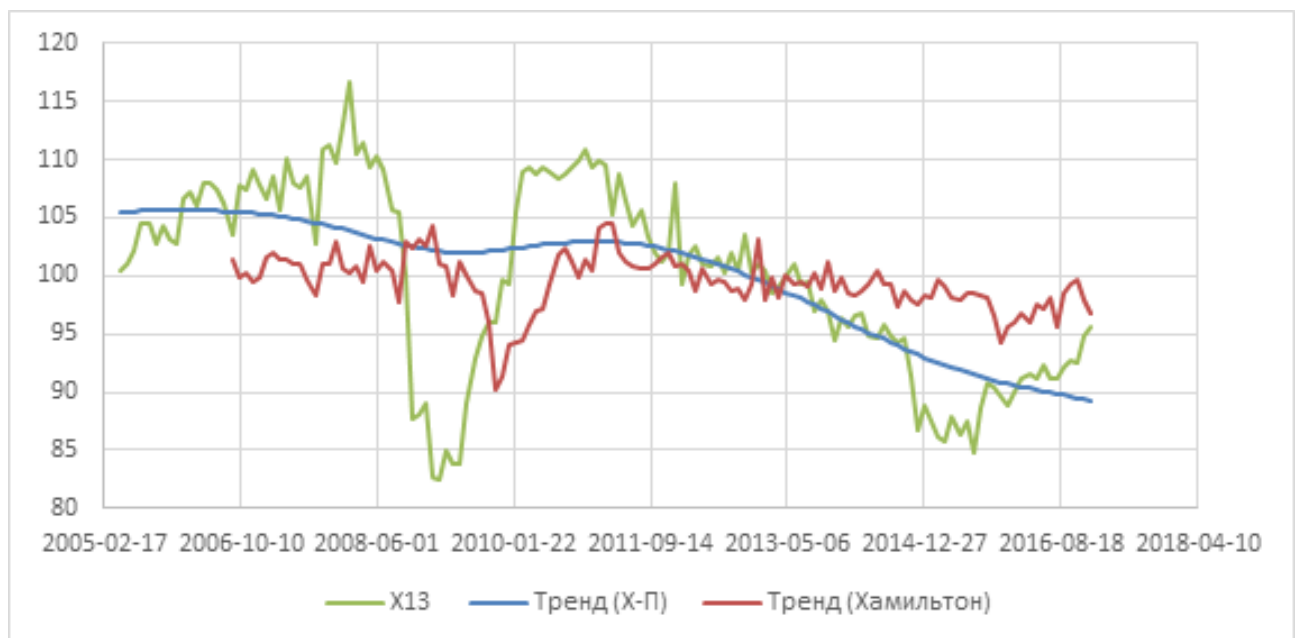


Рисунок 2.3 — Сезонно скорректированный ряд ИЭН и тренды, выделенные фильтрами Ходрика – Прескотта (Х-П) и Хамильтона.

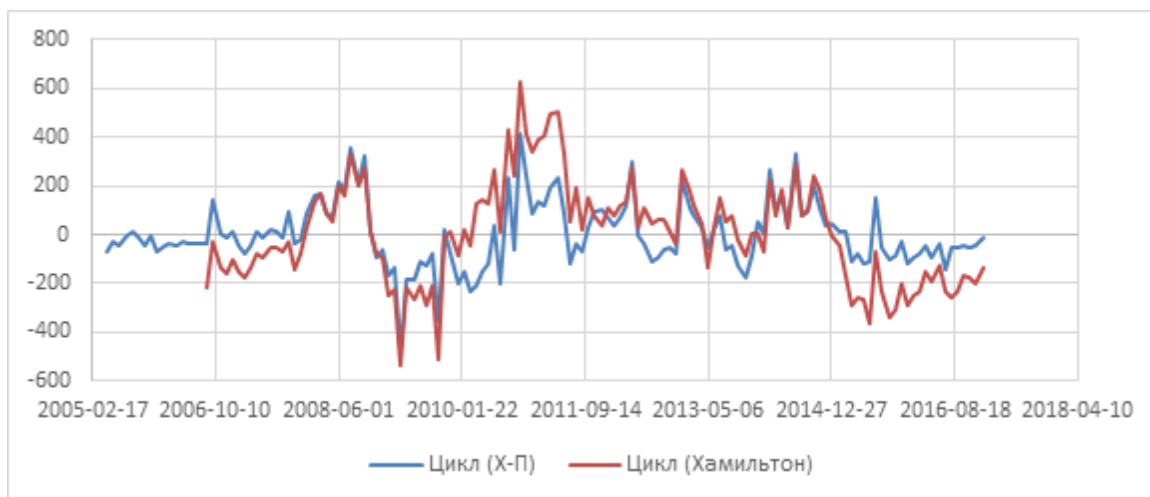


Рисунок 2.4 — Циклические составляющие с шумовой компонентой для реального ВВП.

$$y_{t+12} = \underset{[7.01]}{67.05} + \underset{[2.60]}{0.511}y_t + \underset{[0.54]}{0.125}y_{t-1} + \underset{[0.42]}{0.095}y_{t-2} - \underset{[-2.41]}{0.408}y_{t-3} + \hat{\nu}_{t+h}, \quad \hat{\sigma}^2 = 74.28. \quad (2.2)$$

В уравнениях сохранены все изначально включаемые переменные, вне зависимости от статистической значимости соответствующих коэффициентов, что соответствует тестируемому утверждению Хамильтона об универсальности метода фильтрации. Для фильтра Ходрика – Прескотта не существует аналитической формулы функции тренда.

Остатки моделей (2.1) и (2.2), определяемые как $\hat{\nu}_{t+h} = c_{t+h} + \varepsilon_{t+h}$ и интерпретируемые как циклические составляющие с шумовой компонентой для реального ВВП и ИЭН, изображены на рисунках 2.4 и 2.5 соответственно.

2.1.2 Выделение циклов и анализ поворотных точек

Для выделения долгосрочных циклов с целью их сравнительного анализа к полученным на предыдущем этапе временным рядам циклических составляющих с шумовыми компонентами $\hat{\nu}_t$ применяется фильтр Ходрика – Прескотта с параметром $\lambda = 13.93$. Это позволяет удалить шумовые компоненты и получить гладкие кривые циклов для ВВП и ИЭН, представляемые на рисунках 2.6 и 2.7 соответственно. В таблице 2.1 приведены оценки поворотных точек циклов ВВП и ИЭН полученных с помощью фильтров Ходрика – Прескотта и Хамильтона.

Полученные в рамках настоящего исследования поворотные точки сравнены с соответствующими значениями из [28], полученные с применением фильтра Ходрика – Прескотта. Из таблицы следует, что при использовании фильтра Ходрика – Прескотта в рамках рассматриваемого интервала

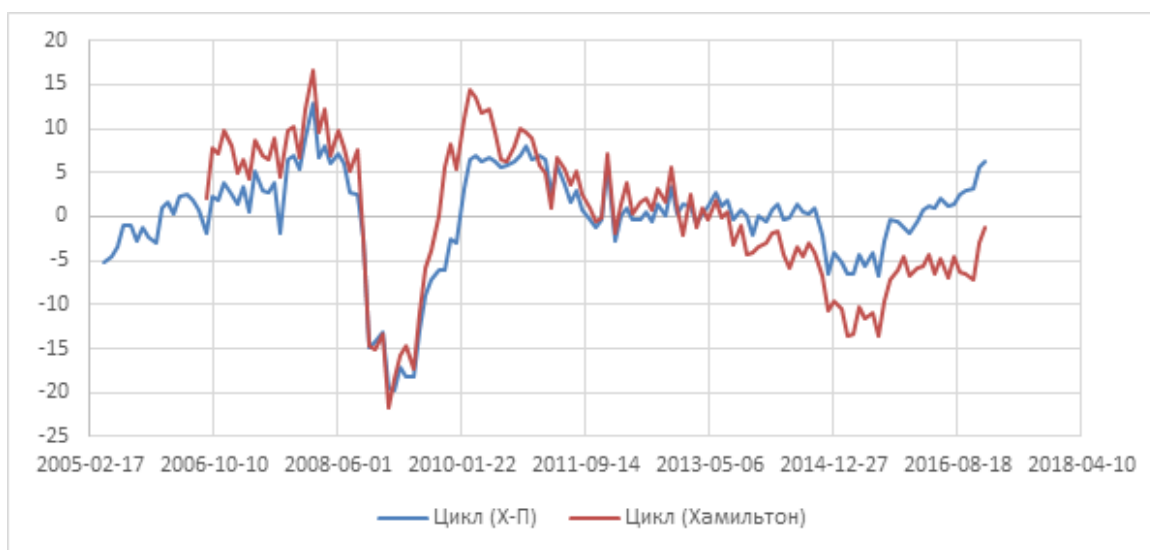


Рисунок 2.5 — Циклические составляющие с шумовой компонентой для ИЭН.

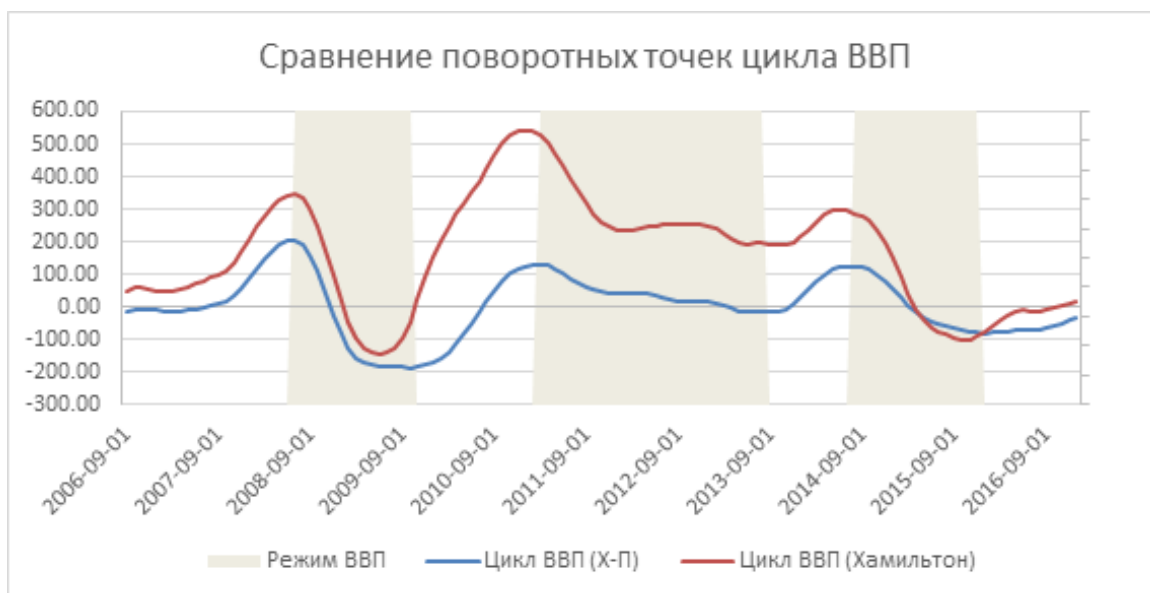


Рисунок 2.6 — Сравнение поворотных точек цикла ВВП. Моменты переключения режима, отображённые цветом фона, являются поворотными точками по Ходрику – Прескотту.

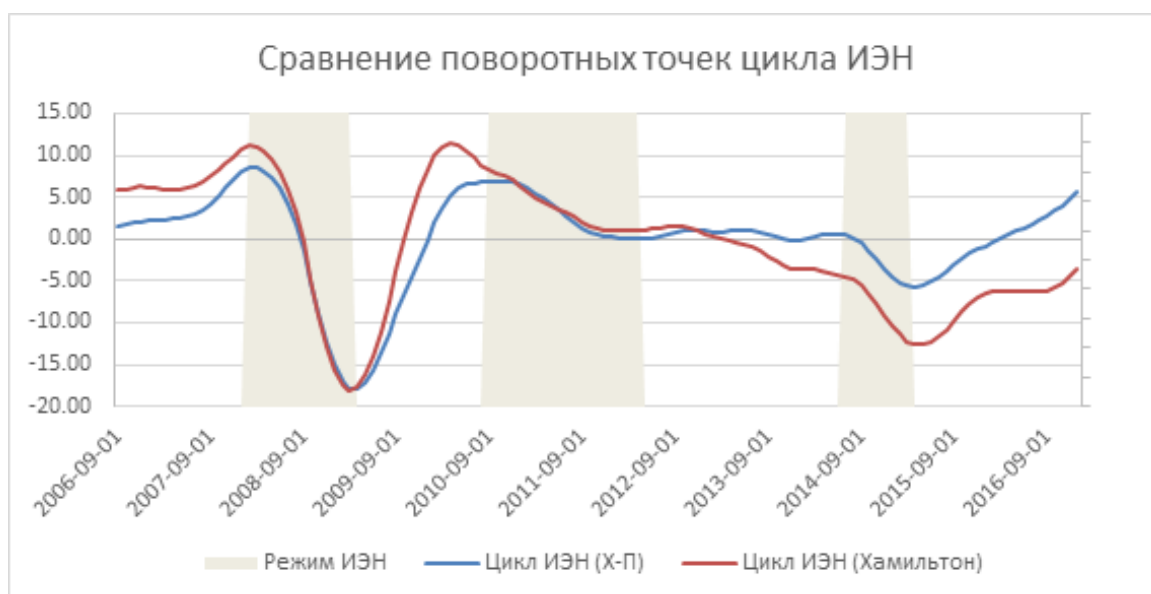


Рисунок 2.7 — Сравнение поворотных точек цикла ВВП. Моменты переключения режима, отображённые цветом фона, являются поворотными точками по Ходрику – Прескотту.

Таблица 2.1 — Сравнение поворотных точек, полученные на основании разных методов фильтрации.

Состояние экономики	Поворотные точки реального ВВП по [28]	ВВП		ИЭН	
		Фильтр Х-П	Фильтр Хамильтона	Фильтр Х-П	Фильтр Хамильтона
Пик	2008.06	2008.06	2008.06	2008.01	2008.01
Дно	2009.09	2009.10	2009.07	2009.03	2009.02
Пик	2011.03	2011.03	2011.01	2010.08	2010.05
Дно	-	-	-	2012.04	2012.01
Пик	-	-	-	-	2012.09
Дно	2013.09	2013.09	2013.09	-	-
Пик	2014.07	2014.07	2014.07	2014.06	-
Дно	2016.01	2016.01	2015.11	2015.03	2015.04

наблюдения с мая 2005 г. по январь 2017 г. поворотными точками являются: июнь 2008 – пик, сентябрь 2009 – дно, март 2011 – пик, сентябрь 2013 – дно, июль 2014 – пик, январь 2016 – дно. Первая фаза экспансии (до июня 2008 г.) связана с чрезвычайно благоприятной внешней экономической конъюнктурой, а последующие (с июня 2008 г. до сентября 2009 г.) фазы замедления роста и спада связаны с воздействием на национальную экономику глобального экономического кризиса. Последующий цикл (сентябрь 2009 – март 2011 – сентябрь 2013) связан с «разогревом» экономики посредством инструментов экономической политики и последовавшим за этим валютным кризисом 2011 г., который привел к замедлению и спаду. Период с сентября 2013 г. по январь 2016 г. отмечен небольшим восстановлением, которое обусловлено некоторым улучшением внешней конъюнктуры и последующими затяжными фазами замедления роста и сжатия, которые были вызваны масштабными изменениями внутренней (ужесточение экономической политики) и внешней (снижение цен на сырьевые товары) среды. В то же время, не было ни одного ложного сигнала о поворотных точках. «Провал» в предсказании двух поворотных точек (сентябрь 2013 г. и июль 2014 г.) можно объяснить высокой неустойчивостью экономической конъюнктуры в этот период и, как следствие, высокой неопределенностью ожиданий экономических субъектов. Таким образом, в подобных ситуациях опережающий индикатор выступает как показатель степени неопределенности будущей экономической конъюнктуры.

Оценки поворотных точек циклов для рассматриваемых временных рядов, полученные с помощью фильтров Ходрика – Прескотта и Хамильтона, в целом достаточно близки. Тем не менее, в контексте задачи выделения экономических циклов можно указать на достаточно частое опережение моментов смены фаз циклов, получаемых с помощью фильтра Хамильтона. На такую особенность данного фильтра указывалось также в [20].

2.1.3 Выводы

Фильтры Хамильтона и Ходрика – Прескотта имеют общий недостаток – наличие априорно задаваемых параметров h и λ соответственно, от которых существенно зависят свойства получаемых трендов и циклических составляющих. Представляется, что фильтр Хамильтона более требователен к длине временных рядов. В условиях относительно короткого времени наблюдения белорусских макроэкономических временных рядов и трудностью априорного задания параметра h , более предпочтительным для рассматриваемых в настоящем исследовании задач является применение фильтра Ходрика – Прескотта. В то же время, в силу отсутствия смещения значений тренда в конечных точках и, соответственно, в остаточной составляющей, использование фильтра Хамильтона может быть предпочтительнее в задачах прогнозирования [A4].

2.2 Анализ циклических изменений и оценка поворотных точек экономических циклов на основе модели с переключением состояний

2.2.1 Результаты построения и применения модели MS-ARX

Для временных рядов, скорректированных по Хамильтону в плане выделения тренда, для моделирования циклической компоненты было построено несколько моделей $MS(L) - AR(p) - X(esi_{-k})$ с параметрами $L \in \{2, 3\}$, $p \in \{0, \dots, 3\}$, $k \in \{-1, \dots, 6\}$ (см. уравнение (1.13)). Для сравнения моделей использовались критерии:

- индивидуальная значимость коэффициентов модели (по t-статистике) во всех режимах,
- критерий Акаике AIC,
- пороговое значение частоты переключения режимов (от 2 до 10 переключений); это условие – проверка на адекватность поведения режимов.

В результате была выбрана модель для GDP со спецификацией $MS(2) - AR(0) - X(ESI_{-4})$. Оцененная модель представлена уравнениями (2.3) и (2.4).

$$\begin{cases} GDP_t = \underset{[-7.827]}{-0.2042} - \underset{[-3.672]}{0.3326} ESI_{t-4} + \nu_t, \nu_t \sim N(0, 0.0429), & l = 0 \\ GDP_t = \underset{[17.33]}{0.4447} - \underset{[-4.076]}{0.2337} ESI_{t-4} + \nu_t, \nu_t \sim N(0, 0.0241), & l = 1 \end{cases} \quad (2.3)$$

$$M = \begin{bmatrix} 0.9787 & 0.0213 \\ 0.0313 & 0.9687 \end{bmatrix}, \quad m_{i,j} = p[i \rightarrow j], \quad (2.4)$$

где $m_{i,i} = p[i \rightarrow i]$ – вероятность остаться в режиме i . Все коэффициенты оказались значимыми на уровне 0.05. Графики предсказания модели (с вероятностью первого режима) представлены на рис. 2.8; остатки модели исследованы на рис. 2.9.

Аналогичным методом была построена модель для годовых темпов роста ВВП и ИЭН¹. Была выбрана модель со спецификацией $MS(2) - AR(0) - X(ESI_{-2})$. Модель для темпов ростов исходных рядов соответствует уравнениям (2.5) и (2.6).

Графики ВВП, модельных значений, режимов, и анализа остатков так же приведены (рис. 2.10, 2.11).

¹от обоих рядов отнята единица для центровки

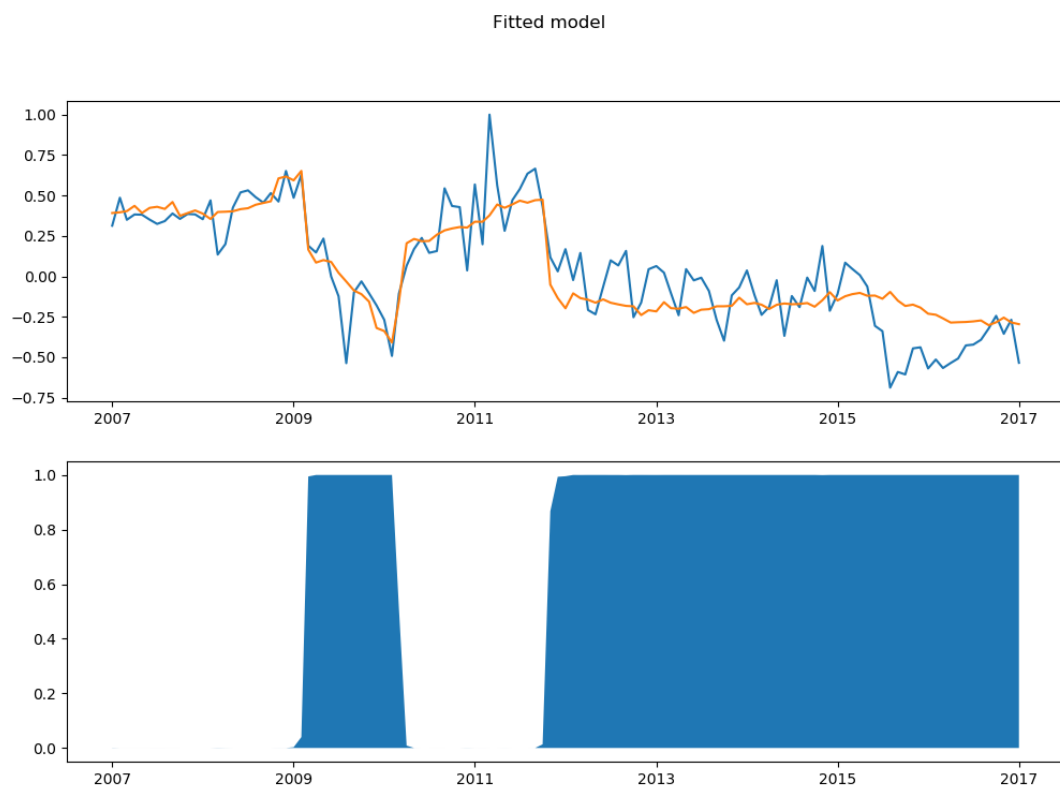


Рисунок 2.8 — Предсказание модели MS(2)-ARX для ВВП по Хамильтону.

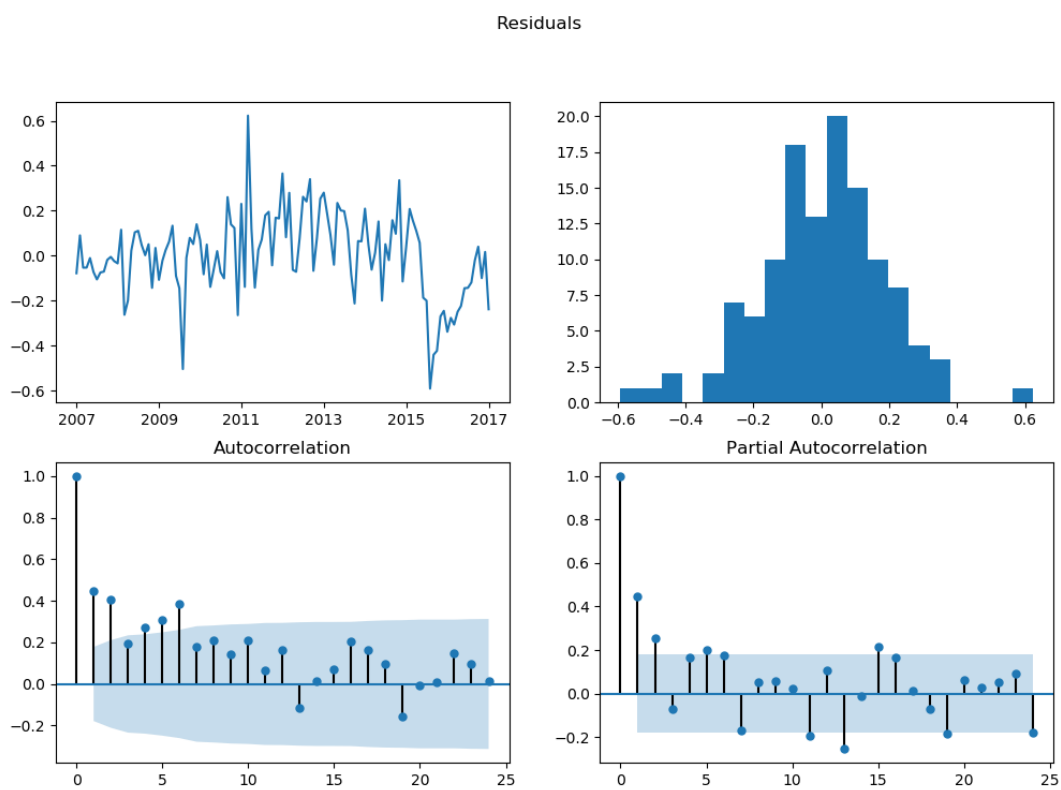


Рисунок 2.9 — Анализ остатков модели для ВВП по Хамильтону.

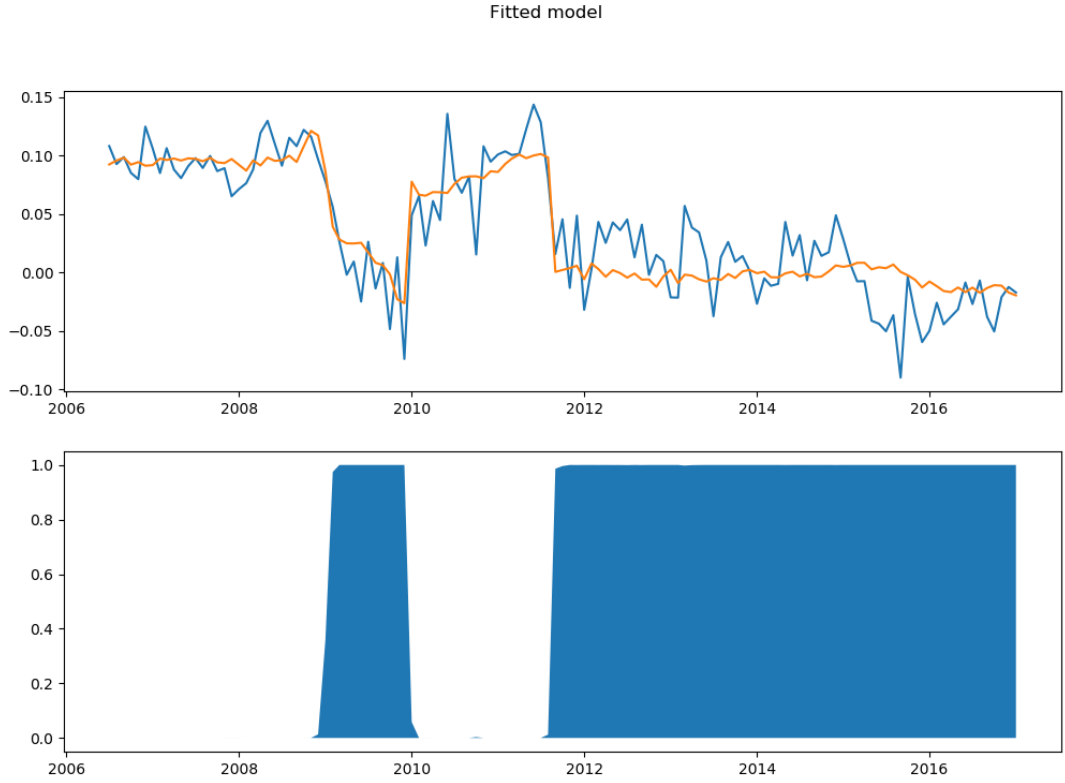


Рисунок 2.10 — Предсказание модели MS(2)-ARX для темпов роста ВВП.

$$\begin{cases} GDP_t = -0.007623 - 0.1432ESI_{t-2} + \nu_t, & l = 0 \\ \quad \quad \quad [-2.060] \quad \quad \quad [-3.333] \\ GDP_t = 0.09731 - 0.1030ESI_{t-2}, & l = 1 \\ \quad \quad \quad [21.78] \quad \quad \quad [-2.785] \\ \nu_t \sim N(0, 0.000786), & l \in \{0, 1\} \end{cases}, \quad (2.5)$$

$$M = \begin{bmatrix} 0.9787 & 0.0213 \\ 0.0313 & 0.9687 \end{bmatrix}, \quad m_{i,j} = p[i \rightarrow j]. \quad (2.6)$$

Обе модели имеют похожие результаты и свойства, несмотря на расхождение по оптимальному опережению (2 и 4). Можно предположить, что фильтр Хамильтона также может являться приемлемым преобразованием для моделирования, однако требуются дальнейшие исследования на различных рядах.

2.2.2 Сравнительный анализ поворотных точек на основе фильтров и модели MS-ARX

В табл. 2.2 приведены оценки поворотных точек разными методами. В качестве «истинных» значений рассматриваются экспертные оценки. Как видно, лучше всего совпадают точки, полученные в ходе двойной фильтрации

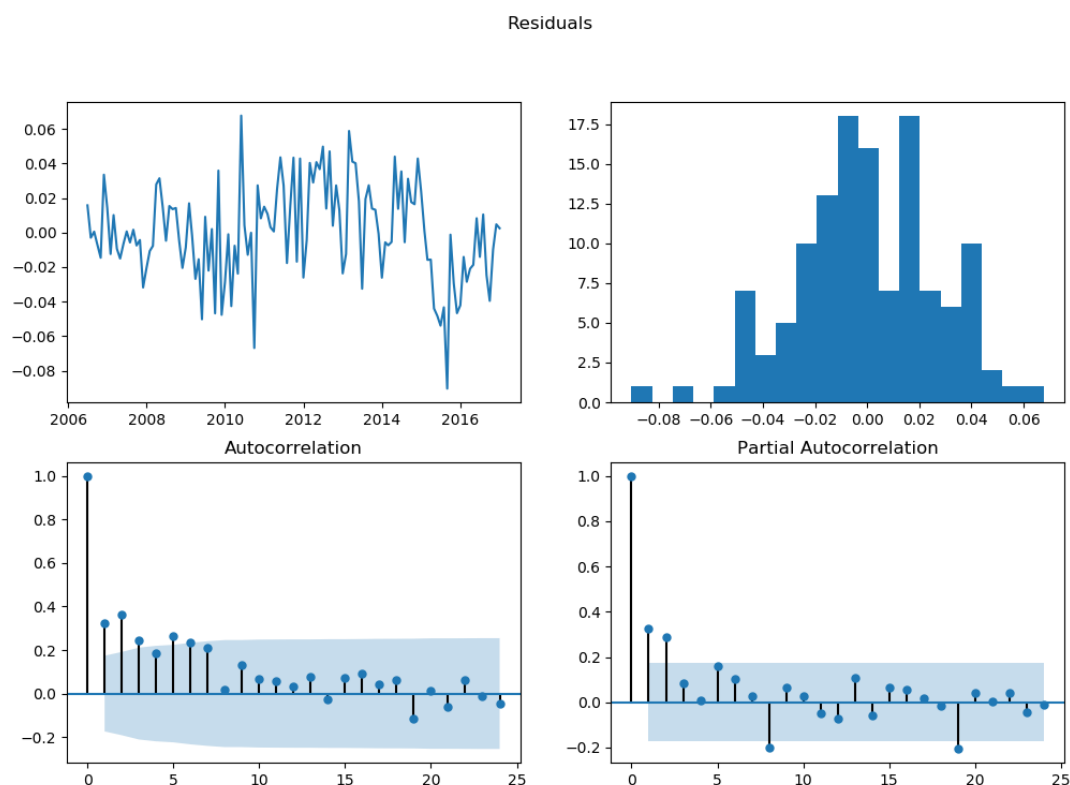


Рисунок 2.11 — Анализ остатков для модели темпов роста ВВП.

(конкретно, для одной точки предсказание 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.

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, чем MS-ARX. В главе 3 настоящей работы описана разработка библиотеки, упрощающая создание такой модели.

Таблица 2.2 — 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	2010.02	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	2011.09	-
Метод	пик 3	дно 3	Замечания	
Экспертные оценки	2014.07	2016.01	Официальные оценки Методика совпадает с офиц.	
Двойной HP filter	2014.07	2016.01		
X13 + Hamilton (сглаж.)	2014.06	2015.11		
MS-VAR для индикаторов доверия	2014.11	-	см. [23]	
MS-ARX для темпов роста ВВП (экзогенная – ИЭН)	-	-		
MS-ARX для Hamilton ВВП (экзогенная – ИЭН)	-	-		

2.3 Прогностическая способность моделей

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

2.3.1 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 при применении для моделирования временных рядов эта процедура имеет серьезные недостатки. Во-первых, если выбрасывать данные из «середины» временного ряда, то становится невозможным предсказание последующих наблюдений. Во-вторых, из-за серийной корреляции оценка ошибки предсказания будет значительно ниже чем действительное значение. 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.

Во избежание 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 оценка метрики (часто – среднеквадратичное или среднее абсолютное отклонение) по $\hat{\varepsilon}_{T-L+1} \dots \hat{\varepsilon}_T$, которая в дальнейшем используется 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.

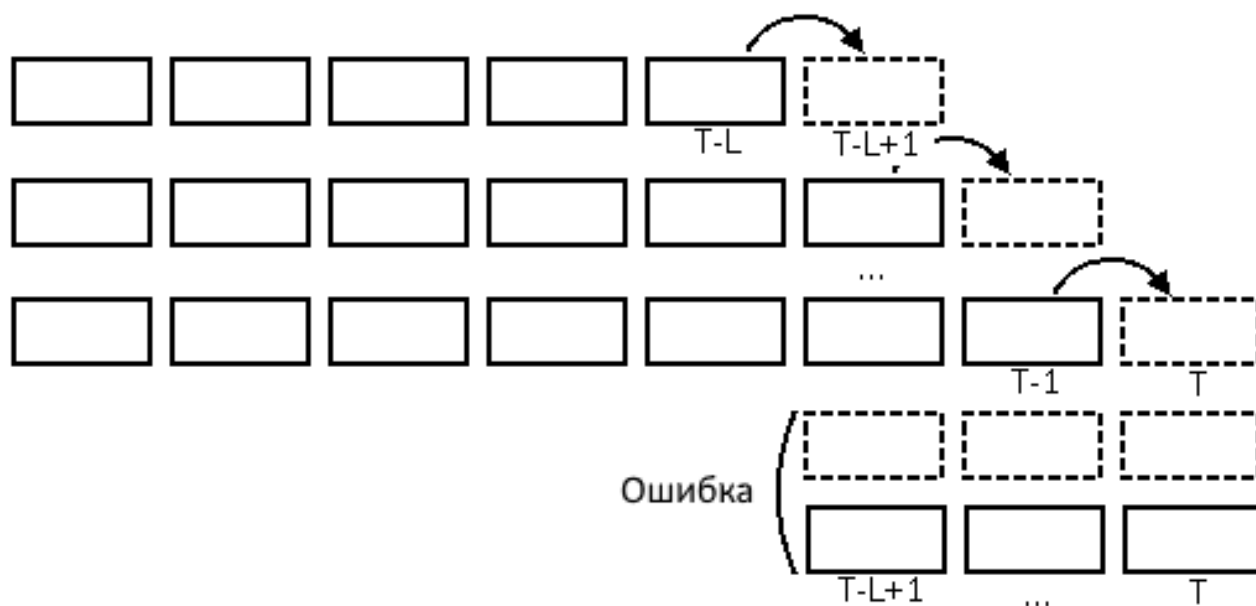


Рисунок 2.12 — Принцип скользящей валидации.

2.3.2 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 [13] и имплементирован в пакете Python **pyramid_arima** [21]). В таблице 2.3 указаны среднеабсолютные (MAE) и среднеквадратичные (RMSE) ошибки прогнозов моделей для ВВП, скорректированные по Хамильтону, и для годовых темпов роста ВВП. При сравнении необходимо учитывать, что у рядов различные масштабы. Ниже представлены графики прогнозов моделей.

Как видно по обоим метрикам и по рис. 2.13 и 2.14, модели $MS(2) - AR(0)X$ сильно уступают подобранным моделям $SARIMAX$. Можно сделать вывод, что модели переключения среднего значения не годятся для предсказания ВВП, а только для классификации прошлых периодов (циклов) и оценивания поворотных точек.

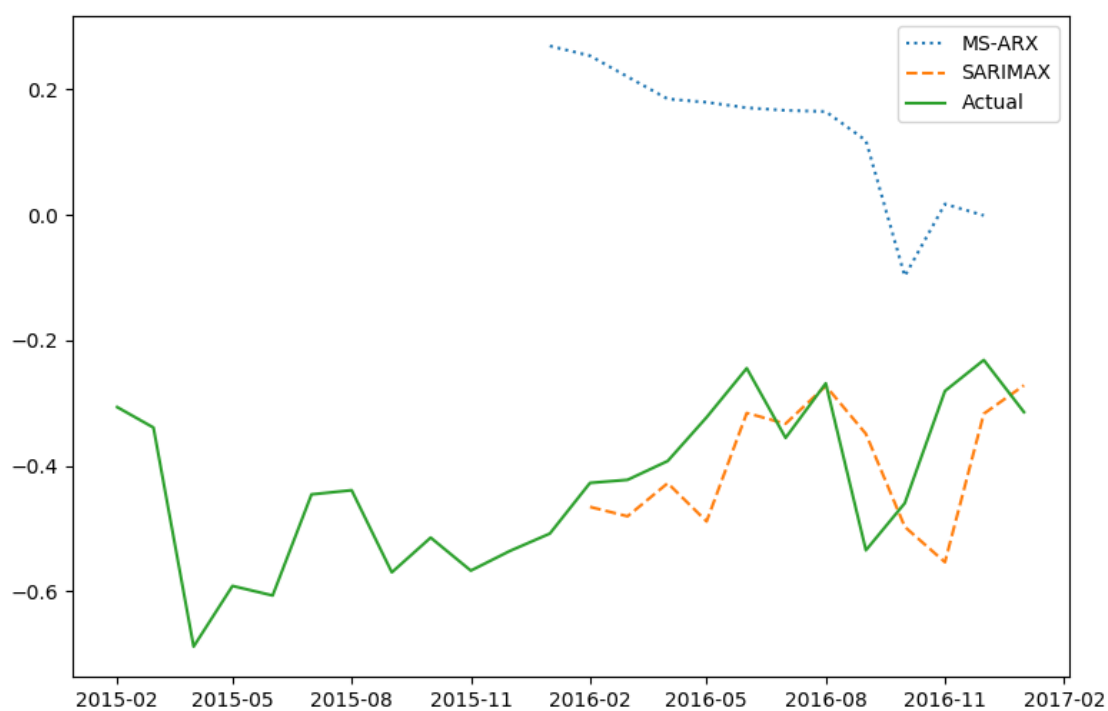


Рисунок 2.13 — Сравнение скользящих прогнозов по Хамильтону.

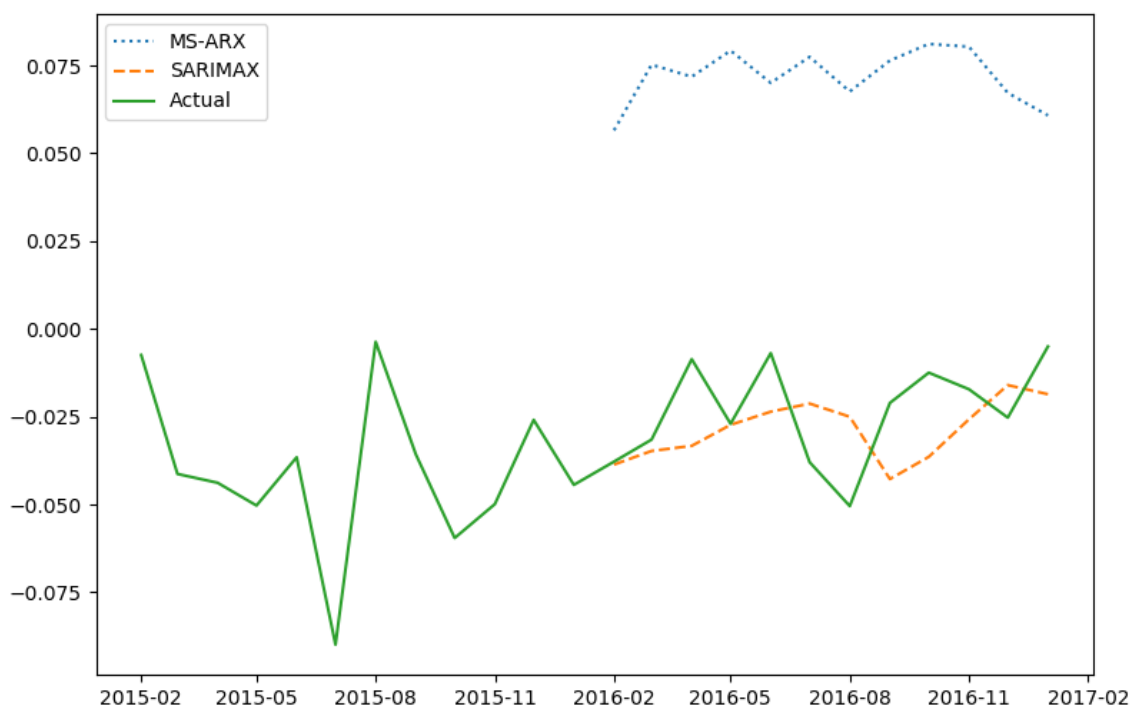


Рисунок 2.14 — Сравнение скользящих прогнозов для темпов роста ВВП.

Таблица 2.3 — Сравнение ошибок прогнозов

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

Глава 3

Разработка библиотеки Time Series for Data Science

Одно из направлений дальнейших исследований – применение векторных моделей переключения состояния (MS-VARX) для отдельных секторов экономики РБ. Для оценивания параметров такой модели уже разработан ЕМ–алгоритм [14], 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 программного пакета.

Реализация алгоритмов оценивания требует значительных усилий, включая работу 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 дисциплине «data science», автором была разработана библиотека «Time Series for Data Science» (коротко – **ts4ds**) на языке программирования Python. Промежуточные результаты проектирования и разработки библиотеки послужили основой выступлением на конференциях [A3, A5].

3.1 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).

3.2 Принципы разработки библиотеки

Для выполнения поставленных задач был необходим фундаментальный подход к проектированию и разработке 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 оптимальную¹ (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

¹Оптимальную модель, либо аппроксимацию оптимальной.

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 нахождение оптимального эстиматора можно рассматривать как алгоритм поиска оптимальной точки (комбинации параметров) в некотором подпространстве.

3.3 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. Эстиматоры (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**[19], 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.е. такой, как в **statsmodels**):

- неудобно сохранять модели компактно, так как нет явного понятия «минимальных данных для воспроизведения»;
- эстиматоры, которые оценивают одинаковые модели (например, МНК и LASSO), должны заново воплотить процедуру предсказания; более того, могут присутствовать 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.

3.4 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 соответствующими 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;

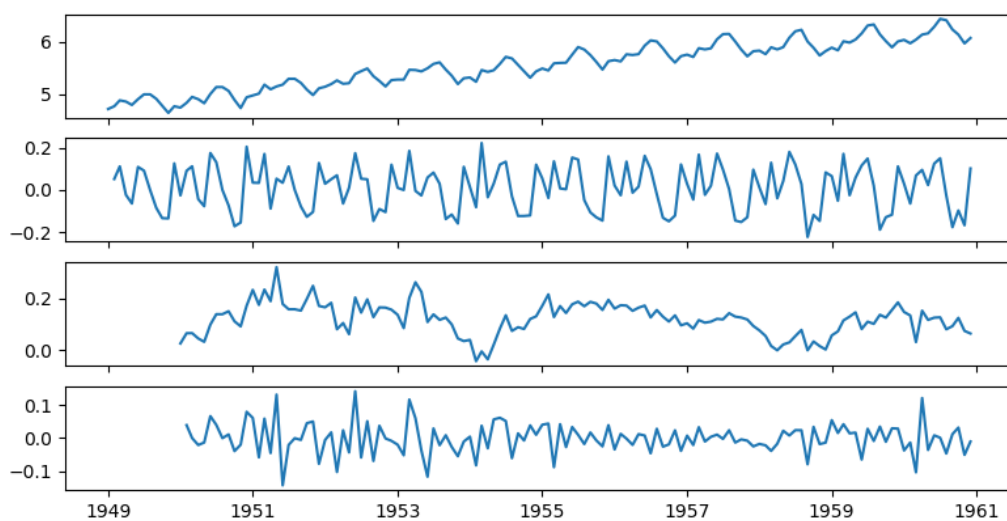


Рисунок 3.1 — График результата выполнения (после различных трансформаций).

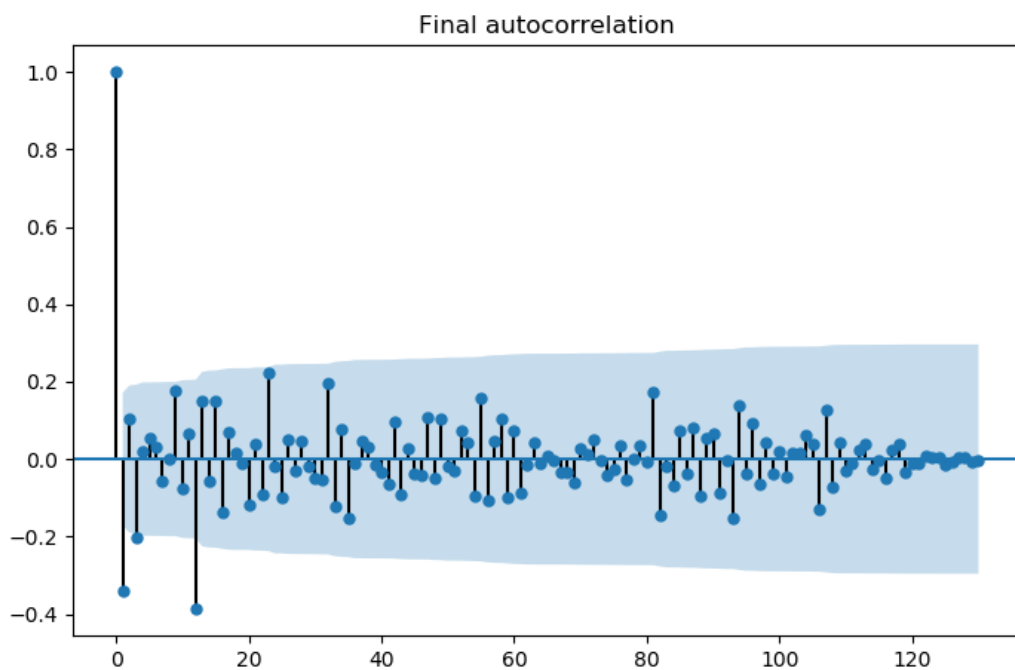


Рисунок 3.2 — График результата выполнения (АКФ после сезонных и первых разностей).

- общая модель дискретного пространства состояний (statespace models);
- модели ARIMAX, SARIMAX.
- модели с независимыми и Марковскими переключениями состояний (семейства IS-VARX, MS-VARX).

Так же реализованы две процедуры автоматического подбора порядков для SARIMAX, так называемое «auto_arima»[21]. 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).

3.5 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*:

$$\begin{cases} y_t = \beta^T x_t + u_t \\ \Phi(L)\phi(L)\Delta_s^Q\Delta^q u_t = \Theta(L)\theta(L)\varepsilon_t \\ \varepsilon_t \sim N(0, \sigma^2) \end{cases} \quad (3.1)$$

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» (statespace model [6]) и предсказывать с помощью фильтра Калмана.

Модель statespace описывается следующими уравнениями:

$$\begin{cases} \alpha_{t+1} = c + T\alpha_t + R\eta_t \\ y_t = d + Z\alpha_t + \varepsilon_t \\ \eta_t \sim N(0, Q) \\ \varepsilon_t \sim N(0, H) \end{cases} \quad (3.2)$$

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 основе для предсказаний в **ts4ds**.

На рис. 3.3 показан пример использования модели *SARIMAX* для прогнозирования на 4 периода вперед для сезонной модели, а также ряд ошибок предсказаний (код для получения этих результатов – в листинге).

```

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()

```

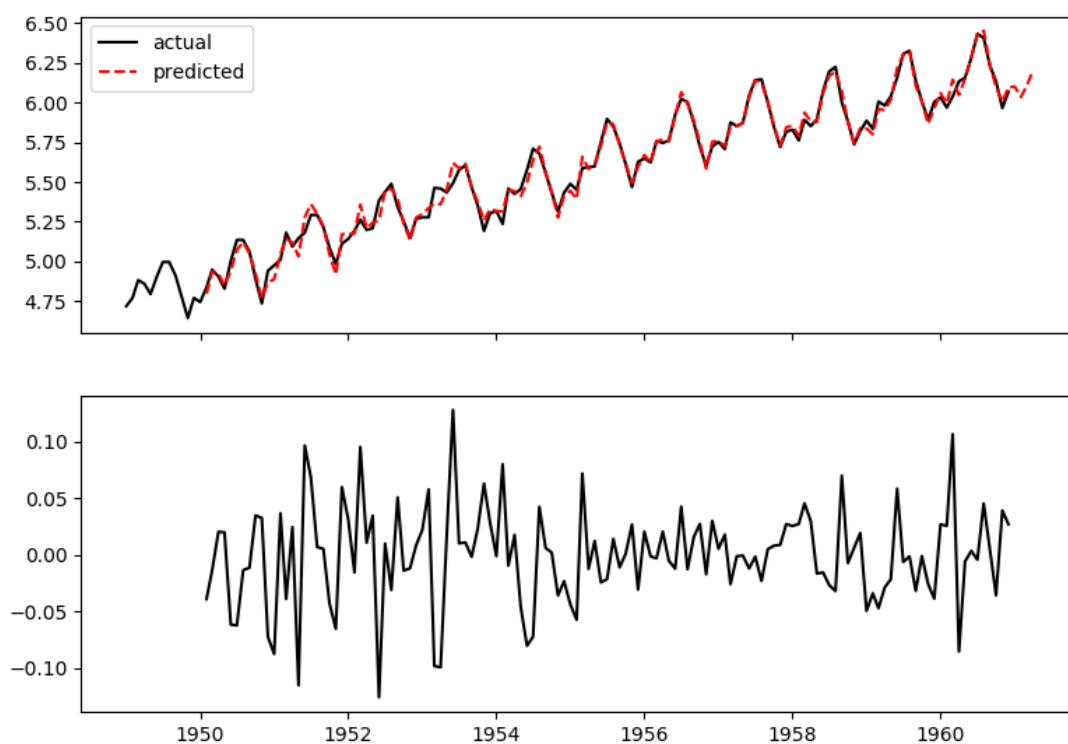



Рисунок 3.3 — Пример работы модели SARIMAX.

Заключение

В работе получены следующие основные результаты:

- исследованы возможности применения фильтров Ходрика – Прескотта и Хамильтона для выделения циклов и оценивания поворотных точек бизнес-цикла белорусской экономики, проведено сравнение получаемых оценок с экспертными оценками;
- сделан вывод, что фильтр Ходрика – Прескотта предпочтительно использовать в задачах классификации и для более коротких временных рядов, а фильтр Хамильтона – в задачах прогнозирования;
- построены модели с переключением состояний типа MS-ARX для анализа бизнес-цикла белорусской экономики (в качестве альтернативы методам, основанные на фильтрах);
- исследованы предиктивные возможности построенных моделей типа MS-ARX и SARIMAX для реального ВВП;
- спроектирована и разработана библиотека программ **ts4ds** на языке Python для анализа временных рядов с элементами машинного обучения, включающая использованные в работе эконометрические модели и методы анализа.

Все решаемые задачи являются новыми и имеют практическую значимость. Задача сравнения фильтра Хамильтона с фильтром Ходрика – Прескотта на данный момент является предметом продолжающихся теоретических и экспериментальных исследований в европейских и американских институтах. Данная работа является первым подобным исследованием на основе белорусских данных; в целом, результаты схожи с результатами упомянутых исследований. Установлено, что модели MS-ARX, рассмотренные для решения задач нахождения поворотных точек и прогнозирования значения ряда ВВП, пока уступают по точности рассмотренным альтернативам. Разработанная библиотека позволит упростить создание и применение более продвинутых моделей с переключением состояния.

Материалы (исходный код и данные) первых двух глав работы размещены в открытом доступе на сайте <https://github.com/NowanIlfideme/PyEconModelling>. Библиотека **ts4ds** используется в компании «ИООО ЭПАМ Системз»; исходный код библиотеки на данный момент еще не опубликован, однако идет процесс предоставления исходного кода в открытый доступ.

СПИСОК ИСТОЧНИКОВ

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. *Harvey A., Jaeger A.* Detrending, stylized facts and the business cycle // Journal of Applied Econometrics. — 1993. — Т. 8. — С. 231—247.
11. *Haubrich J.* How Cyclical is Bank Capital? — February 2018. — URL: <https://ssrn.com/abstract=3124452> (дата обр. 11.04.2018).
12. *Hodrick R., Prescott E.* Postwar U.S. business cycles: An empirical investigation // Journal of Money, Credit and Banking. — 1997. — Т. 29. — С. 1—16.
13. *Hyndman R., Khandakar Y.* Automatic Time Series Forecasting: The forecast Package for R // Journal of Statistical Software. — 2008. — Т. 27, № 3.
14. *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.

15. *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.
16. *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).
17. *OECD.* OECD System of Composite Leading Indicators. — 2012. — URL: <http://www.oecd.org/std/leading-indicators/41629509.pdf> (дата обр. 26.05.2016).
18. *Pederson T.* The Hodrick-Prescott filter, the Slutsky effect, and the distortionary effects of filters // Journal of Economic Dynamics and Control. — 2001. — Т. 25. — С. 1081—1101.
19. *Perktold J., Seabold S., statsmodels-developers.* Python package Statsmodels on statsmodels.org. — 2017. — URL: <http://www.statsmodels.org/stable/index.html>.
20. *Schuler Y.* Detrending and financial cycle facts across G7 countries: Mind a spurious medium term! // TECB Working Paper Series. — 2018.
21. *Smith T.* Python package pyramid-arima on Github/PyPi. — 2017. — URL: <https://github.com/tgsmith61591/pyramid>.
22. *Statistical Research C. for, Methodology.* X-13ARIMA-SEATS Reference Manual / U.S. Census Bureau. — 2016. — 284 p.
23. *Бабахин Е.* Анализ циклических изменений в экономике Республики Беларусь на основе моделей VAR с Марковскими переключениями состояний : Курсовая работа / Бабахин Е. — БГУ, 2015.
24. *Демиденко М., Кузнецов А.* Оценка равновесного реального ВВП: фильтр Ходрика–Прескотта // Банкаўскі веснік. — 2011. — Т. 1, Студзень. — С. 19—26.
25. *Зубарев А., Трунин П.* Определение разрыва выпуска для российской экономики // Российское предпринимательство. — 2016. — Т. 17, № 3. — С. 381—388. — doi: 10.18334/rp.17.3.2225.
26. *Малюгин В.* Методы анализа многомерных эконометрических моделей с неоднородной структурой. Научное издание. — 2014. — 351 с.

27. *Малюгин В.* 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.
28. *Малюгин В., Крук Д.* О построении и применении индекса экономических настроений белорусской экономики // Экономика. Моделирование. Прогнозирование / под ред. М. Кравцов [и др.]. — Минск: НИЭИ Министерства экономики Республики Беларусь, 2017. — Вып. 12. — С. 141—157.
29. 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 Минэкономки Республики Беларусь. 2017. — С. 178—188.
30. *Харин Ю., Малюгин В., Харин А.* Эконометрическое моделирование: учеб. пособие. — Минск : БГУ, 2003. — 313 с.

Список публикаций автора

- A1. *Макаревич А., Малюгин В.* Сравнительный анализ фильтров Ходрика – Прескотта и Хамильтона при оценивании поворотных точек бизнес-цикла белорусской экономики / А.С. Макаревич, В.И. Малюгин // Банкаўскі веснік. – № 8/661. Жнівень 2018. – с. 49-56.
- A2. *Макаревич А., Малюгин В.* Анализ поворотных точек бизнес-цикла белорусской экономики на основе алгоритмов Ходрика – Прескотта и Хамильтона / А.С. Макаревич, В.И. Малюгин // Статистические методы анализа экономики и общества: 9-я международная научно-практическая конференция 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.
- A3. *Makarevich A.* Automating the Process of Construction and Application of Time Series Models by Means of Machine Learning / A. Makarevich // Book of Abstracts : EURO mini-conference Logistics Analytics 2018, Minsk, 18-19 June 2018. – p. 18.
- A4. *Макаревич А.* Сравнительный анализ оценок поворотных точек экономического цикла на основе алгоритмов Ходрика – Прескотта и Хамильтона / А.С. Макаревич, рук. В.И. Малюгин // 74-я научная конференция студентов и аспирантов Белорусского государственного университета. — Белорусский Государственный Университет. 2017. — часть 1, с. 53-57.
- A5. *Макаревич А.* Проектирование библиотеки программ для анализа временных рядов / А.С. Макаревич, рук. В.И. Малюгин // 75-я научная конференция студентов и аспирантов Белорусского государственного университета. — Белорусский Государственный Университет. 2018. (в печати).

Приложения. Научные публикации автора.