МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

АЛГОРИТМ СГЛАЖИВАНИЯ ВРЕМЕННЫХ РЯДОВ

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 248 группы
направления 09.04.03 — Прикладная информатика
механико-математического факультета
Спиридонова Кирилла Александровича

Научный руководитель
доцент, к. ф.-м. н. _____ С. С. Волосивец

Заведующий кафедрой
д. ф.-м. н. _____ С. П. Сидоров

Саратов 2019

ВВЕДЕНИЕ

Актуальность темы исследования. Динамика некоторых экономических процессов в одни и те же периоды времени в разные годы может быть схожей, то есть иметь сезонность. Наличие сезонности во временном ряду может затруднять анализ исследуемого явления или даже приводить к некорректным выводам. А при анализе зависимостей между экономическими явлениями, наличие схожей сезонной компоненты даже между рядами, совершенно между собой не связанными, может привести к ошибочному выводу о наличии между данными процессами устойчивой статистически значимой связи. Таким образом, задача устранения сезонности временного ряда принимает исключительно важное значение.

Первые работы, посвященные алгоритмам сглаживания временных рядов, появились еще в 60-х годах прошлого века. И с тех пор данные вопросы не теряют своей актуальности. Активная работа в этом направлении ведется постоянно. Наиболее значимых результатов в разработке алгоритмов сезонного сглаживания достигли Банк Испании, Бюро переписи населения США, Европейский центральный банк. В Российской Федерации методы сезонной корректировки в своей работе используют многие организации: Росстат, Банк России, Министерство экономического развития, Газпром, Альфа Банк и другие.

Актуальность определила выбор **темы** данной работы: "Алгоритм сглаживания временных рядов".

Целью работы является теоретическая разработка алгоритма для сезонного сглаживания временных рядов, а также его практическая реализация (включая разработку графического интерфейса) на языке программирования R.

Для достижения данной цели в работе решаются следующие **задачи**:

- изучение и систематизация наиболее распространенных алгоритмов сглаживания, в том числе сезонного, временных рядов;
- проведение теоретической разработки алгоритма сезонного сглаживания;
- написание на языке R скрипта, реализующего разработанный алгоритм;

- проверка адекватности работы разработанного алгоритма путем сравнения его с наиболее популярными алгоритмами сезонной корректировки;
- разработка графической оболочки для алгоритма сглаживания с использованием библиотеки Shiny.

Теоретико-методологической основой исследования служат работы Лукашина Ю.П., Льюиса К.Д., Носко В.П., Светунькова И.С., посвященные эконометрике и, в частности анализу и прогнозированию временных рядов, работы Maravall A., Gomez V., Бюро переписи населения США, посвященные сезонному сглаживанию, а также документация языка R и его отдельных пакетов.

Информационной базой исследования являются данные официальной статитстики Федеральной службы государственной статистики Российской Федерации (Росстат).

Практическая значимость - практическая реалзиация разработанного алгоритма совместно с графической оболочкой представляет собой самостоятельный программный комплекс, который может полноценно использоваться для анализа и работы со временными рядами, имеющими сезонную составляющую.

Работа прошла **апробацию** на различных конференциях, в частности, в VI Международной молодежной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2017 года; в XIX Международной Саратовской зимней школе "Современные проблемы теории функций и их приложения посвященной 90-летию со дня рождения академика П. Л. Ульянова, январь 2018 года; на ежегодной студенческой конференции "Актуальные проблемы математики и механики проведенной механико-математическим факультетом СГУ в апреле 2019 года, в секции "Анализ данных".

Выпускная квалификационная работа состоит из введения, трех основных глав, заключения, списка использованных источников и двух приложений.

Основное содержание работы

Введение содержит основные положения: актуальность темы исследования; цель, объект, предмет, задачи исследования; практическую значимость исследования.

Первая глава **"Обзор алгоритмов сезонного сглаживания"** содержит историю развития и обзор алгоритмов сглаживания, в том числе сезонного, временных рядов.

В **параграфе 1.1** рассматриваются простейшие алгоритмы сглаживания временных рядов. Методы, описанные в данном разделе не подразумевают проведения декомпозиции ряда на части и работают с рядом, как с единым целым. Под сглаживанием здесь скорее понимается некоторый способ локального усреднения данных, при котором несистематические компоненты взаимно погашают друг друга.

В **параграфе 1.2** описаны более сложные алгоритмы сглаживания, подразумевающие наличие у ряда некоторой структуры. Методы приводятся в порядке усложнения. Так, методы Брауна и Хольта предполагают наличие у ряда только линейного тренда. Результаты применения данных алгоритмов к рядам с сезонностью приводят к неудовлетворительным результатам. Для учета наличия сезонной компоненты существует модификация метода Хольта, предложенная Винтерсом и впоследствии получившая название метод Хольта-Винтерса. Также показано, что в зависимости от типа аггрегации компонент временного ряда (аддитивный или мультипликативный) применяются разные алгоритмы, учитывающие это различие.

Далее рассмотрены наиболее современные алгоритмы сезонного сглаживания, разработанные в Бюро переписи США и Банке Испании. Отличием данных методов от описанных выше являются более сложные математические методы, использующиеся в процессе сглаживания, а также многие другие дополнительные возможности: заполнение пропущенных значений, учет количества рабочих и выходных дней, учет плавающих праздников и т.д.

Вторая глава **"Алгоритм сезонного сглаживания"** посвящена теоретической разработке алгоритма сезонного сглаживания временных рядов. Раздел начинается с описания проблемы несопоставимости различных данных, публикуемых Росстатом.

В **параграфе 2.1** описывается первичное преобразование данных. Пусть в распоряжении имеется ряд $x_1, \dots x_n$, тогда количество полных лет за которые имеются наблюдения можно определить как

$$T = \left[\frac{n}{p}\right] + \gamma, \quad p = 4, 12,$$

где $[x]$ - целая часть числа x, а

$$\gamma = \begin{cases} 0, & n \mod p = 0; \\ 1, & n \mod p = 1. \end{cases}$$

В зависимости от агрегации основных компонент временного ряда, ряды могут иметь аддитивную либо мультипликативную структуру. Работа с аддитивной моделью кажется более предпочтительной поскольку сезонная волна становится более однородной и не эволюционирует со временем. Поэтому имея в распоряжении временной ряд с мультипликативной компоновкой необходимо перед сглаживанием его логарифмировать, для перехода к аддитивной модели.

Пусть $x_1, \dots, x_n$ - исходный временной ряд с мультипликативной структурой, тогда переход к аддитивной модели $y_1, \dots, y_n$ осуществляется

$$y_t = \ln(x_t + |\alpha|), \text{где}$$

$$\alpha = \begin{cases} \min_t(x_t) - 1, & \text{если } \exists t : x_t \leq 0; \\ 0. \end{cases}$$

Таким образом, на первом этапе проводится линеаризация данных с целью перехода к аддитивной модели

$$y_t = C_t + S_t + I_t.$$

Параграф 2.2 посвящен второму этапу алгоритма сезонного сглаживания, а именно описанию фильтра Ходрика-Прескотта.

Фильтр Ходрика-Прескотта используется для выделения продолжительных тенденций временного ряда, то есть тренда. По своей сути фильтр - это решение задачи оптимизации. С одной стороны, сглаженный ряд не должен сильно отличаться от исходного ряда, то есть необходимо минимизировать сумму квадратов отклонений

$$\sum_{t=1}^{n}(y_t - L_t)^2 \longrightarrow min.$$

С другой стороны, ряд должен быть достаточно гладким и изменяться не слишком резко

$$\sum_{t=2}^{n-1}((L_{t+1} - L_t) - (L_t - L_{t-1}))^2 \longrightarrow min.$$

Таким образом, фильтр представляет собой двухсторонний линейный фильтр, который вычисляет сглаженный ряд L_t временного ряда y_t за счёт минимизации рассеивания элементов ряда L_t вокруг y_t при условии минимума суммы элементов дважды дифференцированного ряда L_t

$$\sum_{t=1}^{n}(y_t - L_t)^2 + \lambda\sum_{t=2}^{n-1}((L_{t+1} - L_t) - (L_t - L_{t-1}))^2 \longrightarrow min.$$

Параметр $\lambda \geq 0$ является мерой гладкости получаемого ряда. При $\lambda = 0$ сглаженный ряд в точности совпадает с исходным. Если же $\lambda \longrightarrow +\infty$, то ряд приближается к линейному тренду.

В **параграфе 2.3** описывается основная часть работы алгоритма сезонного сглаживания.

После применения к y_t фильтра Ходрика-Прескотта получаем первую оценку тренда L_t^1 и, вычитая её из исходного ряда, получим сумму сезонной и нерегулярной компонент

$$S_t + I_t = y_t - L_t^1.$$

Теперь необходимо отделить нерегулярную компоненту от сезонной волны. Для этого необходимо из $S_t + I_t$ сделать группировку сезонных компонент по периодам сезонности. Имеем p распределений сезонных коэффициентов $s_i^1, \dots s_i^T$, где $i \in [1, p]$. При этом, в некоторых наборах коэффициент последнего периода T может отсутствовать.

Каждый из рядов $s_i^1, \dots s_i^T$ сглаживается скользящей медианой, с шириной окна равной трем. Таким образом, на данном этапе имеем медианно сглаженные коэффициенты S_t^1.

Поскольку данное приближение сезонной волны является достаточно грубым, его необходимо дополнить ошибками, ограниченными некоторой величиной. Кроме того, при отсутствии выбросов внутри окна, сглаженная кривая будет иметь "зубчатый" вид, поэтому её необходимо модифицировать. Для этого сформируем ряд ошибок

$$e_t = S_t + I_t - S_t^1$$

и рассчитаем для него среднеквадратическое отклонение σ. Новый ряд ошибок составим из таких членов e_t, которые не превышают 2σ, а вместо тех, которые превышают, используем 2σ. Более формально

$$\widehat{e_t} = \begin{cases} e_t, & e_t \le 2\sigma; \\ 2\sigma, & e_t > 2\sigma. \end{cases}$$

Прибавим к медианно сглаженному ряду S_t^1 полученный ряд $\widehat{e_t}$ и сгладим его скользящим средним третьего порядка тремя разными способами:

1. Окно состоит из текущего и 2 предыдущих наблюдений;
2. Окно состоит из текущего и 2 последующих наблюдений;
3. Окно состоит из текущего, предыдущего и последующего наблюдения.

В конечном счете получим три сглаженных ряда S_t^{21}, S_t^{22} и S_t^{23}. В качестве конечной оценки сезонной волны используем среднее значение, полученных рядов, то есть

$$S_t^2 = \frac{S_t^{21} + S_t^{22} + S_t^{23}}{3}.$$

Затем для каждого из p распределений преобразованных сезонных коэффициентов $\widehat{s}_i^1, \dots \widehat{s}_i^T$, где $i \in [1, p]$, рассчитаем дисперсию $\widehat{\sigma}_i^2$ (именно дисперсию будем использовать в качестве меры волатильности ряда). Для снижения волатильности набора сезонных коэффициентов можно использовать следующий подход: чем выше дисперсия некоторого сезонного периода, тем, в некотором смысле, сильнее его необходимо сглаживать. Мерой силы сглаживания ряда может служить ширина скользящего окна или распределение весов внутри него. Фактически, чем выше волатильность ряда, тем больше наблюдений будет использоваться при сглаживании.

Теперь для определенности зафиксируем некоторые $i \in [1, p]$ и $t \in [1, T]$ и будем рассматривать только один сезонный коэффициент $\widehat{s}_i^t$. Каждый такой коэффициент будет заменяться некоторым сглаженным значением, зависящим от дисперсии набора коэффициентов в состав которых он входит, то есть

$$\widetilde{s}_t^i = \sum_{j=1}^{T} \widehat{w}_{it}^j(\widehat{\sigma}_i^2)\widehat{s}_i^j, \quad \text{где} \quad \sum_{j=1}^{T} \widehat{w}_{it}^j(\widehat{\sigma}_i^2) = 1 \ \ \forall i, t \ \ \text{и} \ \ \widehat{w}_{it}^j(\widehat{\sigma}_i^2) \geq 0 \ \ \forall i, t, j. \tag{1}$$

$$\widehat{w}_{it}^j(\widehat{\sigma}_i^2) = \frac{w_{it}^j(\widehat{\sigma}_i^2)}{\sum\limits_{j=1}^{T} w_{it}^j(\widehat{\sigma}_i^2)}, \quad w_{it}^j(\widehat{\sigma}_i^2) = \frac{1}{\sqrt{2\pi\widehat{\sigma}_i^2}} \exp\left(-\frac{(j-t)^2}{2\widehat{\sigma}_i^2}\right). \tag{2}$$

Зависимость весов сглаживания от дисперсии продиктована условием снижения волатильности ряда: чем выше дисперсия, тем медленнее веса убывают к нулю, и, следовательно, ряд становится более гладким.

Стоит отметить, что при сглаживании ряда с использованием данных весов возникает проблема: наблюдения, находящиеся ближе к концам набора сезонных коэффициентов, в результате сглаживания сильно зависят от своего исходного значения.

Решение проблемы состоит в том, чтобы в каждом из p распределений сезонных коэффициентов для весов каждого сезонного коэффициента $\widehat{s}_i^t$, вес, соответствующий сглаживаемому коэффициенту, заменить на минимальный

вес в данном наборе, то есть

$$\widehat{w}_{it}^{j}(\widehat{\sigma}_i^2) = \min_t(\widehat{w}_{it}^{j}(\widehat{\sigma}_i^2) : j = t) \text{ для } j = t.$$

После этого новые веса снова нормируются, как в первой части формулы 2. Окончательно с использованием новых весов по формуле 1 пересчитываются все сезонные коэффициенты.

Далее будем работать уже не с каждым отдельным сезонным периодом, а с сезонной волной внутри каждого рассматриваемого года. Для фильтрации оставшихся сезонных ошибок, сгладим сезонные составляющие каждого года скользящим средним третьего порядка тремя разными способами, как это было описано выше.

Теперь осталось нормировать сезонную составляющую так, чтобы в случае аддитивной модели сумма сезонных коэффициентов в течение года равнялась 0, а в случае мультипликативной модели их произведение равнялось 1:

- для аддитивной модели из каждого коэффициента внутри одного года вычтем их среднее значение;
- для мультипликативной модели поделим каждый коэффициент внутри одного года на их среднее геометрическое.

Таким образом, в распоряжении имеются нормированные сезонные коэффициенты $\widetilde{S}_t^3$, и сглаженный ряд $(y_t - \widetilde{S}_t^3)$. Окончательную оценку тренда можно получить, применяя к сглаженному ряду фильтр Ходрика-Прескотта.

Третья глава работы **"Практическая реализация алгоритма сезонного сглаживания"** посвящена практике и содержит в себе описание процесса разработки скрипта, реализующего алгоритм сезонного сглаживания, графического интерфейса, а также сравнению работы полученного алгоритма с наиболее популярными методами сезонной корректировки.

В **параграфе 3.1** описываются некоторые аспекты практической реализации скрипта на языке программирования R. Выбор данного программного продукта обусловлен несколькими факторами. Во-первых, в настоящий момент R является одним из самых популярных и самых мощных языков для статистических вычислений и моделирования, при этом его синтаксис

достаточно прост. Во-вторых, вследствие его высокой популярности для R существует большое множество пользовательских библиотек, реализующих дополнительную функциональность и позволяющих решить практически любую задачу в несколько простых действий.

Для разработки скрипта сезонного сглаживания использовались как стандартные функции и процедуры языка R, так и функции, разработанные сторонними пользователями. В частности, использовались функции, входящие в свободно распространяемые пакеты mFilter и psych.

Параграф 3.2 посвящен описанию процесса разработки графического интерфейса с использованием библиотеки Shiny. Описаны преимущества использования данного пакета, структура и логика работы разрабатываемого графического интерфейса.

В **параграфе 3.3** приводится сравнение разработанного алгоритма сезонного сглаживания с алгоритмами TRAMO/SEATS и X12-ARIMA. Для сравнения алгоритмов использованы несколько рядов макроэкономических показателей:

1. индекс потребительских цен на сахар в Саратовской области;
2. оборот розничной торговли в России в текущих ценах;
3. объем инвестиций в основной капитал Саратовской области в текущих ценах;
4. уровень безработицы (по методологии Международной организации труда) по России.

Ряды подобраны таким образом, чтобы протестировать все возможные комбинации работы алгоритма: на квартальных и месячных данных, с аддитивной и мультипликативной компановкой ряда.

Сравнение разработанного алгоритма с TRAMO/SEATS и X12-ARIMA полностью демонстрирует его работоспособность. Во всех случаях сезонная волна выделена абсолютно верно, что подтверждается экономической логикой рассмотренных выше процессов.

ЗАКЛЮЧЕНИЕ

В настоящей работе описан алгоритм сезонного сглаживания. Преимуществом алгоритма является то, что в процессе сглаживания используется медианное сглаживание, позволяющее избавиться от выбросов и большого количества ошибок, которые присутствуют в данных. Кроме того, добавление в сезонную компоненту некоторой стохастической составляющей позволяет учесть её возможную эволюцию во времени. А благодаря использованию фильтра Ходрика-Прескотта полученная трендовая компонента достаточно успешно аппроксимирует устойчивые долгосрочные тенденции исследуемого показателя. Разработанный сглаживающий фильтр с весами на основе нормального распределения позволяет снизить избыточную дисперсию сезонного фактора.

В практической части работы приведена реализация алгоритма на языке программирования R. При разработке скрипта использовались как стандартные процедуры языка R, так и функции, разработанные сторонними пользователями. Также с использованием библиотеки Shiny разработана графическая оболочка для алгоритма, благодаря которой существенно упрощается и ускоряется процедура сезонного сглаживания.

Сглажены несколько реальных рядов макроэкономических показателей. Показатели были отобраны таким образом, чтобы протестировать все варианты работы алгоритма: квартальные и месячные данные, мультипликативная и аддитивная структура временного ряда. Проведено сравнение разработанного алгоритма с популярными и наиболее часто используемыми в настоящий момент методами сезонного сглаживания временных рядов TRAMO/SEATS и X12-ARIMA.

Результаты сравнения свидетельствуют о следующем:

- В целом сезонная составляющая, полученная при помощи разработанного алгоритма, схожа с сезонностью, которую выделяли TRAMO/SEATS и X12-ARIMA: пики сезонной волны приходятся на одни и те же периоды; тенденции, наблюдаемые в течении одного года схожи; амплитуда колебаний и некоторая эволюция волны во времени также непротиворечивы;

- Различие заключается в однородности сезонной волны. TRAMO/SEATS практически во всех случаях показывает более однородный результат, что обусловлено принципом его построения: при сглаживании любого ряда алгоритм старается минимизировать эволюцию сезонности во времени. В тех случаях, где сезонность эволюционировала результат работы разработанного алгоритма ближе к X12-ARIMA.

С одной стороны стабильность сезонного фактора, как у TRAMO/SEATS, выглядит привлекательно, с другой – эволюция экономических процессов под влиянием множества факторов происходит постоянно (в том числе и эволюция сезонной составляющей) и в таком случае требование стабильности сезонной компоненты является достаточно жестким. Результат работы X12-ARIMA в некоторых случаях, напротив, допускает чересчур сильную изменчивость сезонности. Разработанный алгоритм в смысле стабильности сезонного фактора располагается между этими двумя методами сглаживания и может полноценно использоваться для работы с данными.