МИНОБРНАУКИ РОССИИ

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

Кафедра математической экономики

Анализ рядов динамики производства электроэнергии

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 441 группы

направления 09.03.03 Прикладная информатика

механико-математического факультета

Миловой Людмилы Сергеевны

Научный руководитель
доцент, к.ф-м.н, доцент ________________ В.В. Новиков

Зав.кафедрой
д.ф-м.н, профессор ________________ С.И. Дудов

Саратов 2020 г.

Введение. Учитывая тот факт, что экономические условия развития предприятия, отрасли, страны изменяются во времени, необходимо анализировать динамику этих изменений для успешной реализации функций управления. Одним из приемов, которым целесообразно воспользоваться при оценке эффективности будущих управленческих решений, являются методы прогнозирования, основанные на анализе временных рядов, цель которых – предсказать с той или иной степенью надежности будущие события и учесть этот прогноз при планировании тех или иных управленческих решений.

Настоящая работа посвящена изучению теоретических основ некоторых методов выделения сезонной составляющей временного ряда, а также их приложениям к анализу реальных экономических данных – динамики производства электроэнергии в России за период с января 2013 по январь 2020 года.

Основное содержание работы. Работа состоит из двух глав:

1. Теоретические аспекты анализа временных рядов,
    1. Анализ тренда,
    2. Декомпозиция временного ряда,
    3. Анализ сезонности. Коррелограмма,
    4. Метод авторегрессии,
    5. Выделение сезонных колебаний на основе вариационных принципов,
        1. Непрерывный случай,
        2. Оценка точности выделения сезонной волны,

6. Сглаживания сезонной составляющей суммами Фурье - Лагранжа,

2. Анализ сезонности производства электроэнергии и объемов добычи нефти в России.

В первом разделе описаны шаги выделения тренда.

Трендом в математической экономике называют основную тенденцию изменения временного ряда. Временной ряд (динамический) – это совокупность значений какого-либо показателя, изменяемых во времени.

Выделение тренда состоит из таких действий, как сглаживание исходного временного ряда методом простой или взвешенной скользящей средней (используется как первый шаг при анализе тренда), использование моделей кривых роста для изображения функции тренда, а также прогнозирования ряда в будущих периодах. Наличие шума при работе с временными рядами порождает трудности в его анализе, поэтому используют этап «сглаживание», при котором фактические уровни ряда заменяют расчетными, менее подверженным колебаниям. При сглаживании нерегулярные компоненты и периодические компоненты взаимно уничтожаются.

Самый простой метод сглаживания – метод скользящих средних, в котором каждый член ряда заменяется простым или взвешенным средним g (g – длина интервала сглаживания) соседних членов. Взвешенное скользящее среднее – это скользящее среднее, определяющееся формулой взвешенного арифметического среднего. То есть, каждому члену функции присваивается вес, который равен соответствующему члену арифметической прогрессии.

Второй этап анализа тренда состоит из модели кривых роста, позволяющую представить функцию тренда, получить теоретические значения уровней ряда. Наиболее распространенная кривая роста – класс полиномов различных степеней.

Классическим способом оценки параметров полиномиальных моделей является метод наименьших квадратов (МНК), суть которого состоит в нахождении таких оценок параметров a и b, при которых сумма квадратов отклонений экспериментальных данных (y) от теоретических ($\hat{y}_x$) будет минимальна

$$\sum_i \left(y_i - \hat{y}_{x_i}\right)^2 \to \min.$$

Следующий раздел посвящен декомпозиции временного ряда.

При анализе временного ряда, прежде всего, выявляют тренд ряда, с помощью которого определяют основное направление развития события за длительный период времени. Наряду с трендом имеют место периодические колебания: циклические и сезонные. Помимо этих двух колебаний существуют случайные колебания, появляющиеся из-за второстепенных факторов (случайная компонента). Каждый уровень временного ряда может быть представлен в виде суммы этих компонент:

$$Y = T + S + E.$$

Декомпозиция временного ряда – выделение тренда, сезонной, циклической и случайной составляющих ряда. С помощью декомпозиции можно понять структуру ряда, построить его модель. Большинство процессов в экономике имеют сезонные колебания, воздействие которых требуется исключать при дальнейшем изучении данных процессов. В случае, когда тенденция или периодические колебания (а иногда и то, и другое) отсутствуют, уровни ряда являются функцией случайной.

Третий раздел посвящен анализу сезонности и коррелограмме. В этом разделе объясняется, что такое периодическая зависимость, перечислены свойства коэффициента автокорреляции, даны определения таким понятиям как автокорреляционная функция и частная автокорреляционная функция, а также пояснены различия между ними.

Периодическая зависимость – корреляционная зависимость между каждым i-м и $(i-n)$-м элементом временного ряда, где n – лаг (сдвиг, запаздывание). Суть корреляционной зависимости состоит в том, что при изменении значения одной переменной автоматически изменяется значение другой (других).

Свойства коэффициента автокорреляции:

1. Данный коэффициент строится по аналогии с линейным коэффициентом корреляции, а также характеризует тесноту линейной связи текущего и предыдущего уровней ряда. Таким образом, по коэффициенту автокорреляции можно определять существует ли линейная тенденция у данного ряда.
2. Делать выводы о возрастающей или убывающей тенденции ряда по знаку автокорреляции не имеет смысла, так как большинство временных рядов в экономике содержат положительную автокорреляцию уровней, но убывающую тенденцию.

Корреляционную зависимость выражает автокорреляционная функция (ACF) – последовательность корреляций между рядом и сдвинутыми им лагами на 1,2,..,n,.. временных точек. В статистике автокорреляция описывает корреляцию между значениями процесса из определенного диапазона времени. **Определение 1.** Автокорреляционная функция временного ряда y_t

$$r(t,s) = \frac{\mathrm{M}[(y_t - \bar{y}_t)(y_s - \bar{y}_s)]}{\sigma_t \sigma_s},$$

где М – математическое ожидание; σ_t – среднеквадратическое отклонение y_t; $\bar{y}_t$ – среднее значение y_t; t, s–различные моменты времени, $t-s = n$; y_t – уровень временного ряда в момент времени t.

Если n=1, то коэффициент автокорреляции будет являться коэффициентом корреляции между y_t и y_{t-1}. В стационарном ряде, в котором среднее значение и дисперсия постоянны, автокорреляционный коэффициент равен коэффициенту корреляции между y_t и y_{t-2}; y_t и y_{t-3} и т.д.

График зависимости значений автокорреляционной или частной автокорреляционной функции от величины лага называется коррелограммой.

При анализе автокорреляционной функции и коррелограммы возможно найти лаг, при котором связь между текущим и предыдущим уровнями наиболее близкая, то есть выявить структуру ряда. При наиболее высоком коэффициенте автокорреляции первого порядка ряд содержит только тенденцию. Но при наиболее высоком коэффициенте порядка t ряд содержит циклические колебания с периодичность в t моментов времени. Коэффициент автокорреляции уровней и автокорреляционную функцию следует использовать для выявления во временном ряде наличии или отсутствия трендовой компоненты и циклической компоненты.

Следующий раздел посвящен методам авторегрессии.

Модель, в которой расчетные значения уровней ряда определяются как линейная функция от предыдущих исследований, называется авторегрессионной. Авторегрессионные модели очень широко используются для прогнозирования стационарных временных рядов. Главной особенностью данной модели является то, что вероятностное свойство рядов не меняется во времени.

Значимый коэффициент частной автокорреляции при наибольшем лаге n указывает на существование в исследуемом процессе авторегрессии AR n-го порядка, равного величине данного лага

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} + u_t,$$

где $a_1,\ldots,a_n$ – параметры модели; a_0 – постоянная (часто равная нулю); u_t - белый шум.

Благодаря оценке параметров модели получается механизм прогнозирования, основанный на предположении, что возникшая связь между значениями сохранится некоторое время в будущем.

В пятом разделе рассмотрены некоторые вопросы, связанные с учетом сезонных колебаний, а также алгоритмы выделения таких колебаний.

Сезонные колебания – один из примеров наличия цикличности в экономической динамике. Их необходимо учитывать, так как они значительно затрудняют анализ динамики финансовых и экономических показателей в течение года.

В основе непараметрического алгоритма выделения сезонных колебаний лежит использование вариационных принципов. Важное эмпирическое наблюдение состоит в том, что сезонные колебания устойчивы даже при сильных изменениях тренда, поэтому алгоритм предназначен для выделения стационарных (стабильных) сезонных колебаний.

Определние 2. Непрерывная дважды дифференцируемая функция $S_T(t)$ называется волной, а T – периодом, если выполняется два свойства:

$$S_T(t) = S_T(t + kT),$$

где $t \in [t_0, t_1]$, k – номер периода ($k \leq K - 1$).

$$\int_{t_0+kT}^{t_0+(k+1)T} S_T(t)dt = 0,$$

где $t_0 + (k+1)T \leq t_1$, то есть рассматриваются только циклы (сезонные волны), которые не изменяют суммарных значений показателя за период.

Пусть $x(t) = y(t) - S_T(t)$– функция, доставляющая минимум функционалу:

$$\Phi(x) = \int_{t_0}^{t_1} (\dot{x}(t))^2 dt = \int_{t_0}^{t_1} (\dot{y}(t) - \dot{S}_T(t))^2 dt.$$

тренд $x(t)$ обладает наименьшей суммарной кривизной на интервале наблюдения.

С помощью условия экстремума функционала запишем уравнение для волны с граничными условиями:

$$\frac{d^2 S_T(t)}{dt^2} = f(t).$$

Положим, что $t_0 = 0$, а $t_1 = KT$, то есть приходим к точечной задаче для обыкновенных дифференциальных уравнений со значениями в точках $(k–1)T$.

Представим $y(t),\ t \in [0, KT]$, в виде набора функций, определенных на одном периоде $[0, T]$:

$$y_k(t) = y(t) + (k-1)T,$$

где k = 1,..K. Тогда

$$\frac{d^2 S_T(t)}{dt^2} = \frac{1}{K} \sum_{k=1}^{K} \frac{d^2 y_k(t)}{dt^2}, t \in [0, T].$$

Выразив постоянные S_0 и S_1, получим уравнение

$$S_T(t) = S_0 + S_1 t + \frac{1}{K} \int_0^T \sum_{k=1}^{K} y_k(t) dt,\ t \in [0, T].$$

Данное уравнение выставляет начальное значение тренда $x(0) = y(0) - S_T(0)$ так, чтобы первые условия выполнялись на всем интервале наблюдений. Также видно, что без дополнительных предположений о поведении тренда продолжить волну на всю действительную ось нельзя.

Все существенные ошибки, возникающие при обработке экспериментальных временных рядов, делятся на два класса: экспериментальные ошибки и ошибки численных методов.

Все вычислительные ошибки, то есть их абсолютные значения, пропорциональны амплитуде волны («размаху» колебания на периоде).

С учетом регуляризации (методом добавления некоторых дополнительных ограничений) задача выделения сезонной составляющей становится хорошо обусловленной задачей. Остается рассмотреть неустранимую ошибку разностной схемы вычислений – δ_n и ошибку округления – δ_0.

Неустранимая ошибка разностной схемы вычислений

$$\delta_n \cong O(h^2 s^{(\mathrm{IV})}(\xi)).$$

Следовательно, реальная точность выделения «сезонной волны» и тренда определяются длиной ряда и степенью «изрезанности» исходной реализации. Важно то, что четвертые разности также влияют на ошибку выделения тренда, поэтому необходимо их оценивать либо по исходной реализации, либо по более детальному ряду.

Таким образом, неустранимая ошибка разностей схемы вычислений позволяет охарактеризовать длину ряда: ряд «длинный», если устранимая ошибка заведомо меньше требуемой точности выделения тренда, ряд «короткий», если δ_n сравнима с требуемой точностью определения тренда.

В следующем разделе рассматривается непараметрический метод сглаживания сезонной составляющей, основанный на использовании коэффициентов Фурье-Лагранжа периодической функции.

Пусть задана функция $f \in L^2_{2\pi}$, где $L^2_{2\pi}$- класс (пространство) интегрируемых с квадратом функций периода 2π. Ее ряд Фурье имеет вид

$$f(x) \sim \frac{a_0}{2} + \sum_1^\infty (a_k \cos kx + b_k \sin kx),$$

где a_k и b_k - коэффициенты Фурье, равные

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos kt \, dt, \quad k = \overline{0,n};$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\,dt\,, \quad k = \overline{1,n}.$$

Частичная n-я сумма ряда равна

$$S_n(x) = \frac{a_0}{2} + \sum_1^n (a_k \cos kx + b_k \sin kx) = \frac{1}{\pi}\int_0^{2\pi} D_n(t-x) f(t) dt,$$

где

$$D_n(x) = \frac{1}{2} + \sum_1^n \cos kx = \frac{1}{2}\frac{\sin\left(n+\frac{1}{2}\right)x}{\sin\frac{x}{2}}.$$

Если коэффициенты суммы Фурье n – го порядка функции $f(x)$ периода 2π вычислить по формуле прямоугольников, разделяя период на $2n + 1$ равных частей точками

$$x_k = \frac{2\pi k}{2n+1}, \ k = \overline{0, 2n},$$

то получим сумму, замечательную тем, что она есть тригонометрический полином Лагранжа порядка n, интерполирующий f в узлах $x_k = \frac{2\pi k}{2n+1}$, $k = \overline{0, 2n}$.

$$S_n(f,x) = \frac{a_0^{(n)}}{2} + \sum_1^n (a_k^{(n)} \cos kx + b_k^{(n)} \sin kx),$$

$$a_k^{(n)} = \frac{2}{2n+1}\sum_{j=0}^{2n} f(x_j) \cos kx_j\,, \ k = \overline{0,n},$$

$$b_k^{(n)} = \frac{2}{2n+1}\sum_{j=0}^{2n} f(x_j) \sin kx_j\,, \ k = \overline{0,n},$$

Таким образом,

$$f(x_j) = S_n(f, x_j), \ j = \overline{0, 2n}.$$

Числа $a_k^{(n)}$, $b_k^{(n)}$ называются коэффициентами Фурье-Лагранжа функции f . Тригонометрический полином

$$\widehat{m}_N(f,x) = \frac{a_0^{(n)}}{2} + \sum_{k=1}^{N(n)} (a_k^{(n)} \cos kx + b_k^{(n)} \sin kx),$$

где $N(n) \leq n$ – надлежащим образом подобранное число, представляет собой оператор, аппроксимирующий периодическую функцию f. После выделения тренда временного ряда данный оператор можно использовать для выделения периодической (сезонной) составляющей ряда, поскольку он в определенной степени сглаживает случайную составляющую.

Практическая часть работы заключалась в анализе сезонности производства электроэнергии в России (данные взяты с сайта Министерства Энергетики РФ).

Построив и сравнив модели полиномиального тренда первого и четвертого порядка, приходим к выводу, что наилучшим образом описывает исходные данные вторая модель, так как коэффициент детерминации, показывающий процент дисперсии ряда v1 выше данного значения линейного тренда (4% > 2,6%) и для прогнозирования она предпочтительнее.

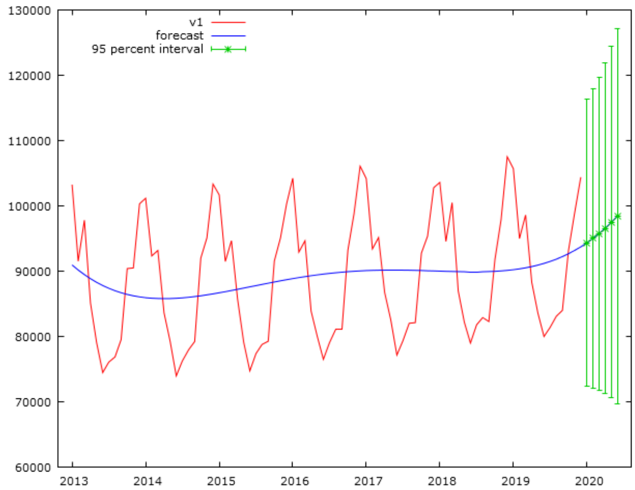

Рис. 1 - Результаты прогнозирования временного ряда v1 с использованием модели полиномиального тренда четвертого порядка

Воспользовавшись процедурами X-12-ARIMA и TRAMO/SEATS для ряда, получим следующие результаты: тренд-циклическая компонента оказывает существенное влияние на формирование уровней ряда, который демонстрирует циклическое поведение, минимальный уровень производства электроэнергии приходится на 2013 год, а максимальный – на 2019 год, временный ряд подвержен эффекту сезонности, влияние случайной составляющей несущественно.

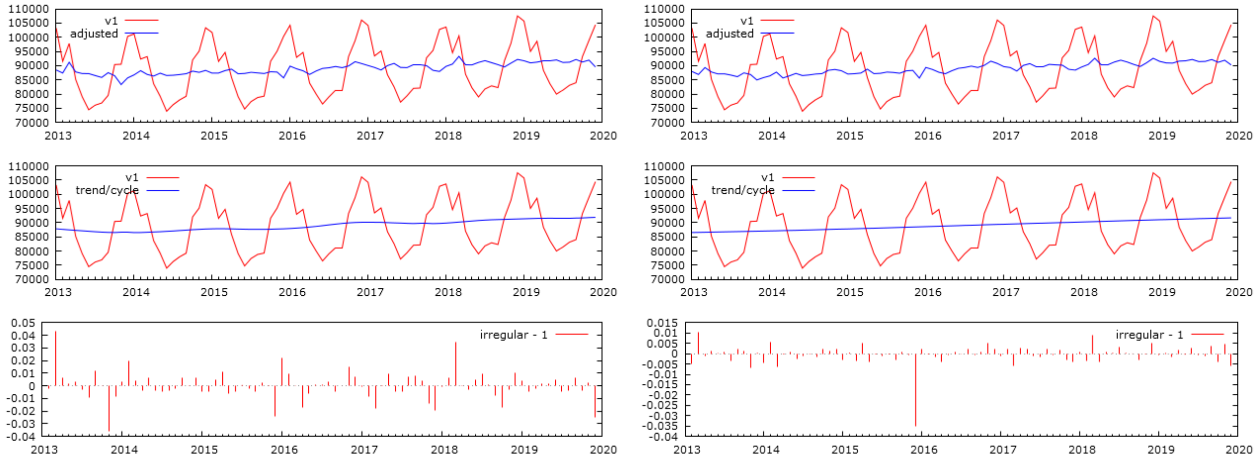

Рис. 2 - Графическое отображение результатов применения процедур X-12-ARIMA и TRAMO/SEATS к временному ряду v1

Заключение. В работе были рассмотрены теоретические основы некоторых стандартных методов декомпозиции и анализа сезонности временного ряда. Кроме того, на основе имеющейся научной периодики рассматривался один из новых подходов к выделению динамической сезонной составляющей, основанный на применении методов вариационного исчисления, а также метод, использующий частичные суммы Фурье-Лагранжа.

Практическая часть была посвящена изучению рядов динамики производства электроэнергии в Российской Федерации. При этом использовались как простые алгоритмы, рассмотренные в первом разделе, так и широко применяемые статистическими организациями методы X-12-ARIMA и TRAMO/SEATS, в основе которых лежит модель ARIMA (AutoregressiveIntegratedMovingAverage – интегрированная модель авторегрессии-скользящего среднего). Кроме того, было выполнено выделение сезонной составляющей на основе сглаживания посредством частичных сумм Фурье-Лагранжа. Все рассмотренные методы показали в целом сопоставимые по точности результаты.