

ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ СЕЗОННЫХ ВРЕМЕННЫХ РЯДОВ ДЛЯ ПРОГНОЗИРОВАНИЯ ДОБЫЧИ УГЛЯ

Спorykhin В.Я., Гонтарев Р.П.

Донецкий национальный технический университет, кафедра АСУ

Abstract

Sporykhin V., Gontarev R. The mathematical model of seasonal time numbers using for coal mining forecasting. The problem of forecasting of seasonal time numbers is considered on the basis of the method of calculation of seasonal waves basing definition of the general tendency. The mathematical forecasting model of a coal mining in Ukraine is received.

Решение проблемы прогнозирования хода и результатов некоторого явления возможно только на основе научного анализа и обобщения предыдущего опыта. Изучение достижений науки и практики дает возможность объяснить и предсказать развитие и поведение какого-либо явления в будущем.

Прогнозируемые явления легче всего описывать временным рядом, который следует рассматривать как сумму детерминированной и случайной компонент. Появление случайной компоненты оценивается с некоторой вероятностью. Детерминированная же компонента выражается некоторой аппроксимирующей функцией, отражающей закономерности развития исследуемого явления.

Таким образом, прогноз некоторого явления по его временному ряду состоит из двух элементов: из прогноза детерминированной компоненты и прогноза случайной компоненты. Разработка первого прогноза не предполагает больших трудностей, если определена основная тенденция развития и возможна ее дальнейшая экстраполяция. Прогноз случайной компоненты более сложен.

Одним из направлений в изучении и анализе временных рядов является выявление сезонности. К сезонным относят явления, которые обнаруживают в своем развитии определенные закономерности, более или менее регулярно повторяющиеся из периода в период.

Статистическое исследование сезонности ставит следующие задачи: численно выразить проявления сезонных колебаний; выявить их силу и характер; вскрыть факторы, вызывающие сезонные колебания.

Известно несколько способов исследования сезонных колебаний: способ простых средних; способ относительных чисел; способ Пирсона; способ расчета сезонных волн, базирующийся на определении тенденции. Наиболее точным, а поэтому и наиболее распространенным является последний способ, который мы и будем рассматривать в данной работе.

Первым шагом в выбранном методе является определение так называемых индексов сезонности, которые характеризуют результаты сравнения фактических уровней данного месяца с уровнями, исчисленными при выявлении основной тенденции (тренда) для этого же месяца. То есть если y_t – фактический уровень временного ряда в определенный момент времени, а \hat{y}_t – уровень основной тенденции в этот же момент времени, то индекс сезонности для этого момента времени вычисляется по формуле (1):

$$I_t = \frac{y_t}{\hat{y}_t} \cdot 100\% \quad (1)$$

Вполне понятно, что индексы по одноименным месяцам за разные годы будут различны. Это объясняется влиянием своеобразных условий каждого года на величину сезонности. Для отражения типичных черт сезонности для каждого месяца пользуются средним индексом за ряд лет:

$$I_k = \frac{1}{n} \sum_{t=1}^n I_t \quad (2)$$

Важным вопросом в рассматриваемом способе является вопрос о методе нахождения тренда данного временного ряда. Методы нахождения основной тенденции развития достаточно разнообразны. Практически же при выявлении тренда ряда, подверженного сезонным изменениям, можно воспользоваться методом наименьших квадратов.

Чтобы полностью учесть влияние несезонных факторов, необходимо, чтобы средняя из индексов сезонности была равна 100. Но так как это редко бывает, то следует производить выравнивание индексов сезонности. Если средняя индексов сезонности I_{cp} , то для выравнивания индексы пересчитываются по следующей формуле:

$$I_k = \frac{I_k}{I_{cp}} \cdot 100 \quad (3)$$

Для временных рядов с ярко выраженной сезонностью расхождение между соответствующими индексами сезонности до их выравнивания и после невелико. Это может быть объяснено, с одной стороны, относительно небольшой колеблемостью по одноименным периодам, а с другой – повторением колебаний из года в год.

Определив влияние сезонного фактора, можно использовать найденные закономерности для прогнозирования дальнейшего развития изучаемого процесса. Как известно, сезонные временные ряды можно разложить на следующие компоненты:

$$x_t = u_t + v_t + \varepsilon_t \quad (4)$$

где u_t, v_t, ε_t – составляющие этого ряда, соответственно: тенденция, сезонные колебания, случайные колебания.

Тенденция отражает общее изменение ряда за длительный промежуток времени: постоянный подъем или постоянное снижение.

Сезонные колебания – это более или менее регулярные изменения временного ряда, возникающие с наступлением данного времени года и повторяющиеся с небольшими отклонениями из года в год.

Случайные колебания вызываются внешними случайными причинами, влияние которых сказывается на уровнях ряда, искажая тенденцию, а также сезонные колебания.

Определив все три составляющие временного ряда, можно попытаться использовать найденные закономерности для экстраполяции их на перспективу. При этом можно оценить значение постоянной компоненты или тренда и краткосрочных сезонных колебаний. Случайные колебания можно оценить только вероятностным путем.

При экстраполяции общих закономерностей на будущий период допускается определенная ошибка: чем продолжительнее период предсказания, тем вероятнее сделать большую ошибку.

В общем виде модель прогноза на любой месяц по каждому из исследуемых показателей буде выглядеть следующим образом:

$$y_t = I_k \hat{y}_t \pm \varepsilon_t, \quad (5)$$

где y_t – фактический уровень ряда в момент времени t ;

I_k – средний индекс сезонности k -го месяца;

\hat{y}_t – оценка уровня ряда в момент времени t ;

ε_t – случайная величина.

Случайную величину ε_t точно определить нельзя. Можно только с определенной вероятностью утверждать, что вычисленные по моделям, которые получены на основании (4), оценки показателя будут отличаться от истинной на величину $t \frac{\sigma_{\varepsilon_t}}{\sqrt{n}}$, где t – число, показывающее, во сколько раз отличается средняя величина от своего отклонения при определенной вероятности, а σ_{ε_t} – среднеквадратическое отклонение случайной величины ε_t .

Применим выше изложенный метод для построения модели прогноза добычи угля в Украине. Имеются данные о добыче угля в Украине за период с октября 1996 года по март 2001 года. Представим имеющиеся данные в виде временного ряда и отразим их в виде графика (ломаной), который приведем на рисунке 1.

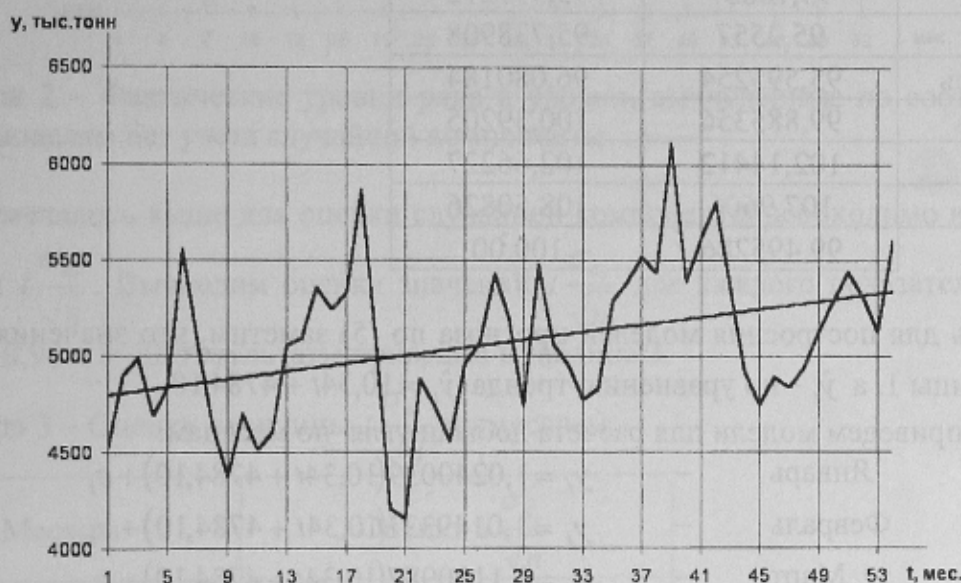


Рисунок 1 – Временной ряд добычи угля и его линейный тренд

Из рисунка 1 видно, что процесс добычи угля имеет колебательный характер. Причем выявляется следующая тенденция: в начале года добыча весьма резко возрастает, затем к середине года начинает убывать, после чего вновь возрастает. Отсюда следует вывод о сезонном характере данного процесса.

Наиболее удобно в качестве основной тенденции воспользоваться линейной моделью $\hat{y}_t = at + b$. Воспользовавшись методом наименьших квадратов определим коэффициенты этой модели: $a = 10,34$, $b = 4784,10$. Таким образом, основная тенденция запишется как $\hat{y}_t = 10,34t + 4784,10$. Положительный коэффициент a говорит о возрастающем характере основной тенденции. Приведем на рисунке 1 график тренда вместе с уровнями временного ряда.

Теперь рассчитаем показатели сезонности по исходным данным по формуле (1) и усредним эти индексы по каждому месяцу с помощью формулы (2). Результаты приведем в таблице 1.

Как говорилось выше для того, чтобы полностью учесть влияние несезонных факторов, необходимо, чтобы средняя из индексов сезонности была равна 100. Для этого необходимо выровнять индексы сезонности по формуле (3). Запишем среднюю индексов в нижней строке таблицы 1 и рассчитаем выровненные индексы в этой же таблице в 3 столбце.

Таблица 1 - Индексы сезонности добычи угля в Украине

Месяц	Индексы сезонности	Выровненные индексы сезонности
Январь	101,88356	102,40039
Февраль	100,98106	101,49331
Март	111,34504	111,90987
Апрель	102,07344	102,59123
Май	92,613132	93,082934
Июнь	89,09736	89,549328
Июль	95,10881	95,591272
Август	95,2557	95,738908
Сентябрь	95,595254	96,080184
Октябрь	99,885356	100,39205
Ноябрь	102,14412	102,66227
Декабрь	107,9606	108,50826
	99,495286	100,00

Теперь для построения моделей прогноза по (5) заметим, что значения I_k определяются из таблицы 1, а \hat{y}_t – по уравнению тренда $\hat{y}_t = 10,34t + 4784,10$.

Ниже приведем модели для расчета добычи угля по месяцам:

$$\text{Январь} \quad - \quad y_t = 1,0240039(10,34t + 4784,10) + \varepsilon_t \quad (6)$$

$$\text{Февраль} \quad - \quad y_t = 1,0149331(10,34t + 4784,10) + \varepsilon_t \quad (7)$$

$$\text{Март} \quad - \quad y_t = 1,1190987(10,34t + 4784,10) + \varepsilon_t \quad (8)$$

$$\text{Апрель} \quad - \quad y_t = 1,0259123(10,34t + 4784,10) + \varepsilon_t \quad (9)$$

$$\text{Май} \quad - \quad y_t = 0,93082934(10,34t + 4784,10) + \varepsilon_t \quad (10)$$

$$\text{Июнь} \quad - \quad y_t = 0,89549328(10,34t + 4784,10) + \varepsilon_t \quad (11)$$

$$\text{Июль} \quad - \quad y_t = 0,95591272(10,34t + 4784,10) + \varepsilon_t \quad (12)$$

$$\text{Август} \quad - \quad y_t = 0,95738908(10,34t + 4784,10) + \varepsilon_t \quad (13)$$

$$\text{Сентябрь} \quad - \quad y_t = 0,96080184(10,34t + 4784,10) + \varepsilon_t \quad (14)$$

$$\text{Октябрь} \quad - \quad y_t = 1,0039205(10,34t + 4784,10) + \varepsilon_t \quad (15)$$

$$\text{Ноябрь} \quad - \quad y_t = 1,0266227(10,34t + 4784,10) + \varepsilon_t \quad (16)$$

$$\text{Декабрь} \quad - \quad y_t = 1,0850826(10,34t + 4784,10) + \varepsilon_t \quad (17)$$

Приведем на рисунке 2 фактические уровни временного ряда добычи угля и уровни ряда, построенные по выше записанным моделям, без учета случайной величины ε_t . На этом рисунке сплошной линией показаны фактические уровни ряда, а штрих пунктирной – уровни, рассчитанные по моделям.

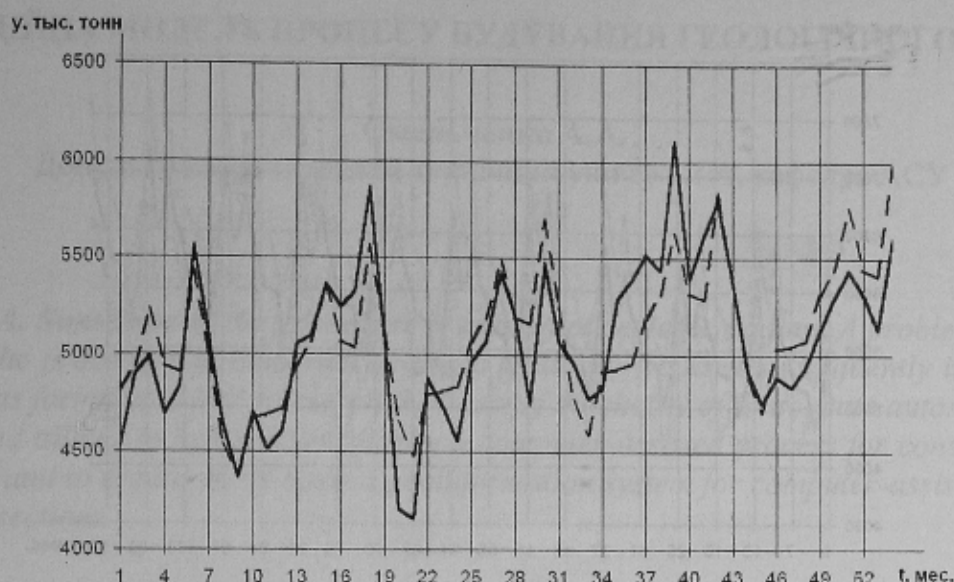


Рисунок 2 – Фактические уровни ряда и уровни, вычисленные по соответствующим прогнозным моделям без учета случайной компоненты

Как отмечалось выше для оценки случайной компоненты необходимо воспользоваться величиной $t \frac{\sigma_{\varepsilon_t}}{\sqrt{n}}$. Вычислим оценки значений $t \frac{\sigma_{\varepsilon_t}}{\sqrt{n}}$ для каждого показателя с вероятностью, равной 0,95 и $t=20$. Результаты приведем в таблице 3.

Таблица 3 – Оценки величины $t \frac{\sigma_{\varepsilon_t}}{\sqrt{n}}$ по месяцам

Месяцы	Оценки $t \frac{\sigma_{\varepsilon_t}}{\sqrt{n}}$
Январь	290,70744
Февраль	576,31937
Март	465,87843
Апрель	188,34172
Май	470,06272
Июнь	355,11113
Июль	176,4668
Август	439,34787
Сентябрь	515,26338
Октябрь	425,75177
Ноябрь	194,33925
Декабрь	536,99132

И теперь пользуясь моделями (6) – (17) выполним прогноз добычи угля в Украине до декабря 2010 года. Приведем на рисунке 3 графическую интерпретацию прогноза, причем покажем также верхнюю и нижнюю границы прогноза, вычисленные при помощи оценки

величины $t \frac{\sigma_{\varepsilon_t}}{\sqrt{n}}$.

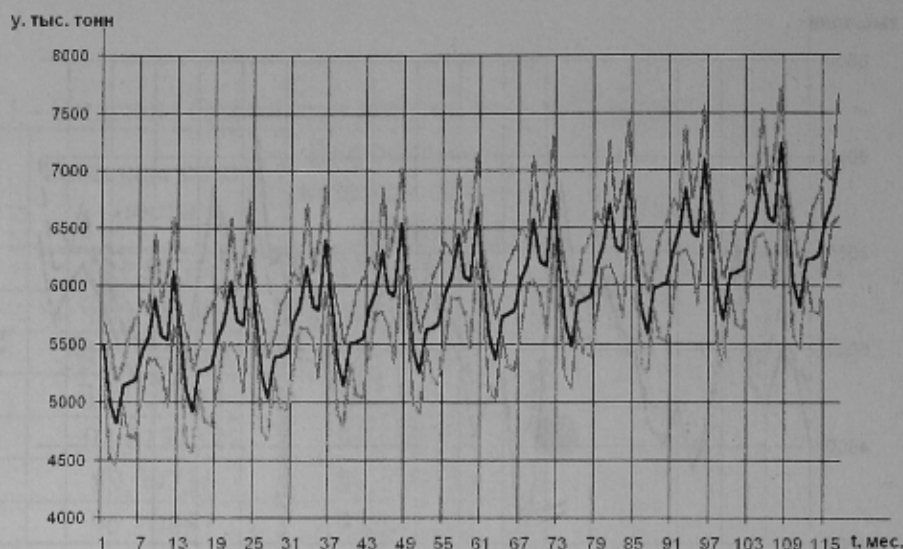


Рисунок 3 – Прогнозный временной ряд добычи угля с нижней и верхней границей

Также приведем в таблице 4 величины добычи угля за год, среднемесячные величины, а также их верхнюю и нижнюю границы.

Таблица 4 – Прогнозные значения добычи угля в Украине

Годы	Прогноз		Верхняя граница		Нижняя граница	
	тыс. т/год	Среднемес. тыс.т	тыс. т/год	Среднемес. тыс.т	тыс. т/год	Среднемес. тыс.т
2002	66034,15	5502,846	70668,73	5889,061	61399,57	5116,631
2003	67523,54	5626,962	72158,13	6013,177	62888,96	5240,747
2004	69012,94	5751,078	73647,52	6137,293	64378,36	5364,863
2005	70502,33	5875,194	75136,91	6261,409	65867,75	5488,979
2006	71991,72	5999,31	76626,3	6385,525	67357,14	5613,095
2007	73481,11	6123,426	78115,69	6509,641	68846,53	5737,211
2008	74970,5	6247,542	79605,09	6633,757	70335,92	5861,327
2009	76459,9	6371,658	81094,48	6757,873	71825,32	5985,443
2010	77949,29	6495,774	82583,87	6881,989	73314,71	6109,559

Как видно из прогноза добычи угля до 2010 года, то существовавшая с конца 1996 по начало 2001 года тенденция роста сохраняется и на прогнозируемый период, соответственно с некоторыми колебаниями, обусловленными сезонностью. Если в 2000 году добыча угля за год составила 62815 тыс. тонн, то в 2010 году можно ожидать 77949,29 тыс. тонн, а с учетом верхней и нижней границы прогноза, угледобыча за 2010 год должна колебаться в пределах 73314,71 – 82583,87 тыс. тонн.

Однако, как отмечалось выше, для продолжительных периодов прогноза высока вероятность ошибки. Так мы вправе считать наш период прогноза продолжительным, так как он более чем в два раза превышает период предыстории. Следовательно, полностью полагаться на прогноз сделанный к 2010 году не следует, так как сделанный прогноз почти на 10 лет опирается на предысторию только лишь за 4,5 года. Для получения более точного прогноза необходимо иметь большую предысторию.

Литература

1. Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. – М.: Статистика, 1973.
2. Хенан Э. Анализ временных рядов. Пер. с англ. – М.: Наука, 1964.