

ОЦЕНКА ЧАСТОТНЫХ ПАРАМЕТРОВ МОДЕЛИ ТЕЙЛА-ВЕЙДЖА В ЗАДАЧАХ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ

В. М. Вартамян

Доктор технических наук, профессор, заведующий кафедрой

Кафедра экономики*

Контактный тел.: (057) 707-46-05

E-mail: vartanyan_vm@ukr.net

Ю. А. Романенков

Кандидат технических наук, доцент

Кафедра менеджмента*

Контактный тел.: (057) 707-46-02

E-mail: k602@d6.khai.edu

В. Ю. Кащева

Старший преподаватель

Кафедра финансов*

Контактный тел.: (057) 707-46-04

*Национальный аэрокосмический университет им. Н.Е. Жуковского «Харьковский авиационный институт»
ул. Чкалова, 17, г. Харьков, Украина, 61070

Розглянуто задачу оцінки параметрів адитивних сезонних моделей для коротких часових рядів. Дано рекомендації щодо спільного застосування адитивної сезонної моделі експоненціального згладжування і методу «Гусениця-SSA», ефективність яких підтверджена прикладом

Ключові слова: прогнозна модель, сезонність, експоненціальне згладжування, тренд, гармонійна складова

Рассмотрена задача оценки параметров аддитивных сезонных моделей для коротких временных рядов. Даны рекомендации по совместному применению аддитивной сезонной модели экспоненциального сглаживания и метода «Гусеница-SSA», эффективность которых подтверждена примером

Ключевые слова: прогнозная модель, сезонность, экспоненциальное сглаживание, тренд, гармоническая составляющая

The problem of an estimation of parameters of additive seasonal models for short time series data is considered. Recommendations about combination of additive seasonal model of exponential smoothing and a Singular spectrum analysis method (SSA) are made. The efficiency is confirmed by an example

Keywords: Forecast model, seasonal prevalence, exponential smoothing, a trend, a harmonic component

1. Введение

Современное состояние проблемы прогнозирования можно охарактеризовать несколькими ключевыми аспектами:

1) современная наука располагает обширным кругом хорошо формализованных методов, позволяющих решать подавляющее большинство практических задач [1];

2) все методы прогнозирования имеют свои ограничения и допущения, и абсолютно универсальных среди них нет и в обозримом будущем не предвидится [2];

3) исследователь при решении практических задач обладает возможностью использовать несколько методов прогнозирования (порой это число даже избыточно) и иногда выбор конкретного метода имеет большее значение, чем его особенности;

4) существует круг задач, которые в силу своих объективных свойств (малая длина, наличие пропусков, высокий уровень зашумленности и т.п.) не могут

быть корректно решены с помощью современных методов прогнозирования без предварительной обработки данных [1, 2].

К упомянутым задачам можно отнести и те, которые находятся «на стыке» областей применения некоторых методов. В этом случае при попытке последовательного применения нескольких методов прогнозирования возможно получение кардинально разных результатов, которые не удовлетворят исследователя. При этом проверка выбранной методики на прошлых данных (так называемый ретроспективный анализ) может давать и удовлетворительные результаты.

2. Постановка проблемы

Рассмотрим класс временных рядов, описывающих процессы, имеющих сезонную динамику. Отметим, что класс таких рядов гораздо шире, чем это кажется. Дело в том, что понятие «сезон» в прогнозировании

применим к любым систематическим колебаниям. Кроме того, цикл колебаний может существенно отличаться от предполагаемых величин (год, месяц, неделя, сутки, час и т.д.). И если удастся выявить величину цикла этих колебаний, то такой временной ряд можно использовать для прогнозирования с использованием аддитивных и мультипликативных моделей [1, 3].

Суть этих моделей состоит в представлении исходного ряда в виде суммы или произведения трех составляющих:

$$F = T + S + E \tag{1}$$

или

$$F = T \cdot S \cdot E, \tag{2}$$

где F – прогнозируемое значение; T – тренд; S – сезонная компонента; E – ошибка прогноза.

На практике отличить аддитивную модель от мультипликативной можно по величине сезонной вариации. Аддитивной модели, в отличие от мультипликативной, присуща практически постоянная сезонная вариация. Отметим, что мультипликативная модель зачастую сводится к аддитивной путем логарифмирования ряда [1].

Сезонная компонента S предполагается периодической с периодом l

$$S(t) = S(t+l), \tag{3}$$

причем для однозначности параметризации модели обычно предполагают, что для аддитивной модели выполняется условие

$$S(1) + S(2) + \dots + S(l) = 0, \tag{4}$$

а для мультипликативной модели выполняется условие

$$S(1) \cdot S(2) \cdot \dots \cdot S(l) = 1. \tag{5}$$

Модификацией метода экспоненциального сглаживания для сезонных рядов являются методы Холта-Уинтерса и Тейла-Вейджа. В качестве модели ряда используется его представление в виде комбинации линейного тренда с сезонной составляющей, наложенной либо мультипликативно (модель Холта-Уинтерса), либо аддитивно (модель Тейла-Вейджа). Предполагается, что коэффициенты тренда и сезонная составляющая могут медленно меняться во времени. В соответствии с этим вычислительный процесс устроен как адаптивная процедура, управляемая тремя параметрами адаптации (один параметр – адаптация уровня, второй – угла наклона, третий – коэффициентов сезонности). В ходе вычислений строится сглаженный ряд, представляющий собой в каждый момент времени t прогноз по данным до момента $(t-1)$ включительно.

Трудности в применении упомянутых методов возникают в случаях, когда неизвестно точное значение периода сезонности, либо ряд содержит менее двух полных периодов. Следует отметить, что при анализе динамики социально-экономических параметров (таких как продажи, спрос и т.п.), нередки случаи, когда реаль-

ное значение периодов сезонности может существенно отличаться от предполагаемых.

Таким образом, **возникает задача оценки периода сезонности** для корректного применения методов экспоненциального сглаживания. Ошибка в оценке этого параметра всего на 8,3% (доля месяца в годовом периоде) чревата получением неверных значений коэффициентов сезонности, и может привести к ошибочному прогнозу.

3. Подход к решению

Будем рассматривать ряд, длина которого L больше одного, но менее двух предполагаемых периодов сезонности:

$$1 \leq L \leq 2l. \tag{6}$$

Для оценки периода сезонности будем использовать метод «Гусеница-SSA», который позволяет разложить исходный ряд на гармонические составляющие. Относительно малая длина ряда является естественным ограничением метода, однако опыт практического использования показал, что в рядах с малым уровнем шума выявить частоту основной гармоники удастся при длине ряда, сравнимой с ее полным периодом.

Период главной гармоники в общем случае может не совпадать с периодом сезонности в силу того, что не всякая сезонная составляющая является гармонической. Однако полученная таким способом оценка выглядит более обоснованной, чем предположения о значении периода сезонности на уровне «здорового смысла».

Кроме частоты основной гармоники, «Гусеница-SSA» позволяет выявить также медленно меняющуюся составляющую (тренд), информацию о которой целесообразно использовать совместно с частотной в аддитивной модели экспоненциального сглаживания.

Предлагается следующая процедура построения прогнозной модели (рис.1.) [5].

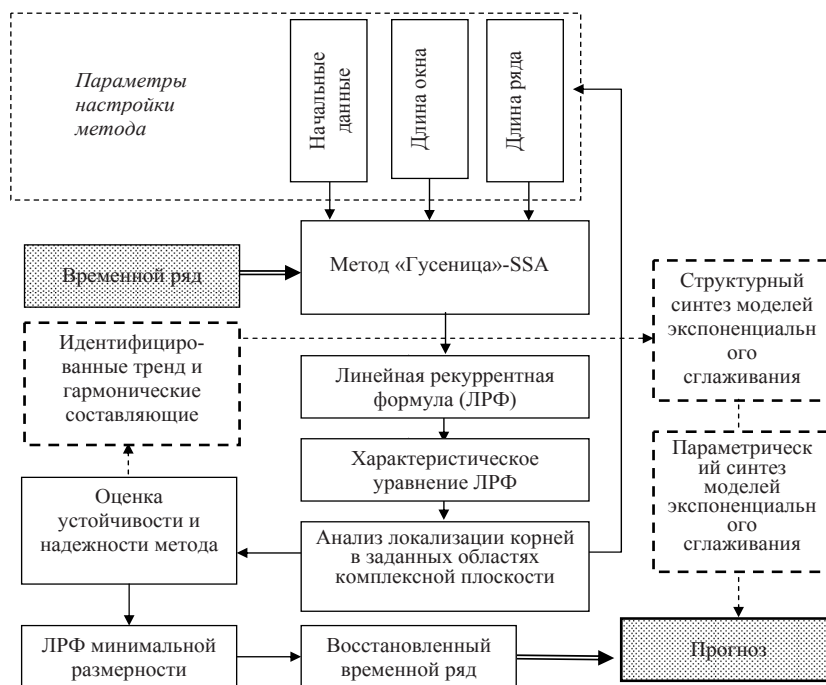


Рис. 1. Процедура структурно-параметрического синтеза рекуррентных моделей динамических процессов методом SSA

Этап 1. Определение тренда. Может быть использован любой из многочисленных подходов выделения тренда. Отметим, что использование полиномиальной формы тренда (например, с помощью стандартных средств MS Excel), хотя и позволяет описать динамику внутри ряда, может привести к существенному искажению итогового прогноза. В случае если физическая природа процесса заранее неизвестна, целесообразным представляется использование либо простейшего вида тренда (линейного или экспоненциального), либо аппарата метода «Гусеница-SSA» для выявления медленно меняющейся составляющей.

Этап 2. Определение сезонной компоненты. Вычитая из фактических значений ряда значения тренда, определяются величины сезонной компоненты.

Этап 3. Определение периода сезонности. С помощью метода «Гусеница-SSA» анализируется исходный ряд и ряд сезонных компонент. Изменяя внутренний настроенный параметр метода - длину окна, можно выявить значение частоты главной гармоники ряда. Отметим, что при выборе разных типов тренда, анализ сезонной компоненты может приводить к разным результатам. Поэтому необходимо следить за корректностью преобразований исходного ряда.

Частоту основной гармоники предлагается определять путем анализа расположения корней характеристического полинома линейной рекуррентной формулы, генерируемой методом «Гусеница-SSA».

Этап 4. Корректировка сезонной составляющей таким образом, чтобы их сумма за период была равна нулю (для аддитивной модели).

Этап 5. Расчет ошибки модели как разности между фактическими значениями и значениями модели.

Этап 6. Построение итогового прогноза с использованием экспоненциального сглаживания, что позволит учесть возможное изменение тенденций, на основе которых построена модель.

4. Пример

Рассмотрим модельный ряд из 28 точек, состоящий из суммы трех гармоник разных частот, экспоненциального тренда и случайной помехи, максимальный уровень которой равен десятипроцентному уровню полезного сигнала.

Воспользовавшись стандартными средствами MS Excel, найдем линейный тренд исходного ряда (рис. 2).

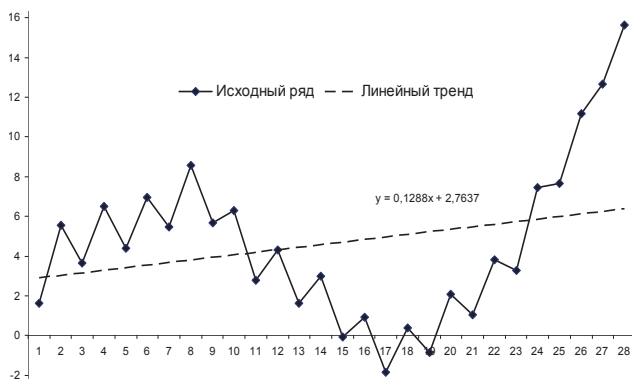


Рис. 2. Линейный тренд исходного ряда

Вычитая из значений исходного ряда значения тренда, найдем сезонную составляющую (рис. 3).

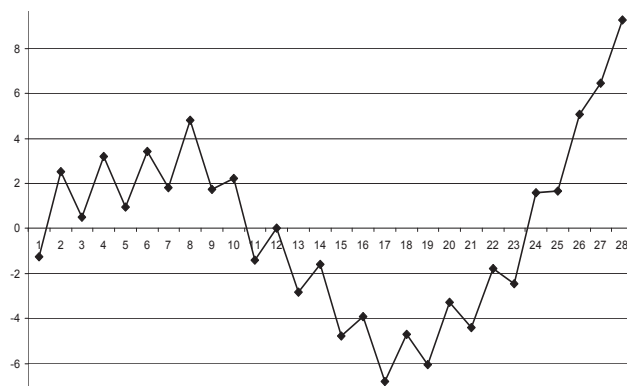


Рис. 3. Сезонная составляющая ряда при выделении линейного тренда

Как видно из рис. 3, сезонная составляющая ряда носит гармонический характер, ряд содержит чуть больше одного периода и последние значения ряда (примерно с 25 точки) явно не повторяют общей тенденции.

Период сезонности по рисунку можно оценить лишь приблизительно, предположительно он равен 23-24.

Разделим ряд на периоды по 24 точки и определим значения сезонных компонент в этих точках. Добавим к тренду поправку в -0,5706, полученную после корректировки сезонной составляющей. Его уравнение примет вид:

$$T = 0,1288t + 2,7637 - 0,5706 = 0,1288t + 2,1931. \quad (7)$$

Определим ошибку прогнозной модели.

t	Значение ряда	1-й сезон	2-й сезон	Среднее	Сезонная компонента	Значение модели	Отклонение
1	2	3	4	5	6	7	8
1	1,6389	-1,2536	1,6605	0,2034	0,7741	3,0959	-1,457
2	5,5425	2,5212	5,0606	3,7909	4,3615	6,8122	-1,2697
3	3,6448	0,4947	6,4435	3,4691	4,0397	6,6192	-2,9744
4	6,4915	3,2126	9,2756	6,2441	6,8148	9,5230	-3,0315
5	4,3762	0,9685	-	0,9685	1,5392	4,3762	0
6	6,9413	3,4048	-	3,4048	3,9755	6,9413	0
7	5,4784	1,8131	-	1,8131	2,3838	5,4784	0
8	8,5885	4,7944	-	4,7944	5,3651	8,5885	0
9	5,6670	1,7441	-	1,7441	2,3147	5,6670	0
10	6,2906	2,2389	-	2,2389	2,8096	6,2906	0
11	2,7770	-1,4035	-	-1,4035	-0,8328	2,7770	0
12	4,3322	0,0229	-	0,0229	0,5935	4,3322	0
13	1,6182	-2,8199	-	-2,8199	-2,2492	1,6182	0
14	2,9722	-1,5947	-	-1,5947	-1,0241	2,9722	0
15	-0,0717	-4,7674	-	-4,7674	-4,1967	-0,0717	0
16	0,9040	-3,9205	-	-3,9205	-3,3498	0,9040	0
17	-1,8460	-6,7993	-	-6,7993	-6,2287	-1,8460	0
18	0,3631	-4,7190	-	-4,7190	-4,1483	0,3631	0
19	-0,8422	-6,0531	-	-6,0531	-5,4825	-0,8422	0
20	2,0675	-3,2722	-	-3,2722	-2,7016	2,0675	0
21	1,0485	-4,4201	-	-4,4201	-3,8494	1,0485	0
22	3,8240	-1,7733	-	-1,7733	-1,2027	3,8240	0

1	2	3	4	5	6	7	8
23	3,2798	-2,4463	–	-2,4463	-1,8757	3,2798	0
24	7,4544	1,5995	–	1,5995	2,1701	7,4544	0
25	7,6442	Сумма		-13,6956	0	6,1871	1,4571
26	11,1731					9,9034	1,2697
27	12,6848					9,7104	2,9744
28	15,6457					12,614	3,0317
Сумма квадратов						952,3744	43,5456

Среднеквадратическая ошибка модели равна 0,041474 или 4,15%. Однако уровень этого показателя говорит лишь о том, что модель строится на данных, вмещающих в себя чуть больше одного периода (рис. 4). Если бы точек было 24, ошибка была бы нулевой. Именно по этой причине, в некоторых программных продуктах не допустима реализация описываемого метода при объеме данных меньшем, чем за два полных сезона. Ошибка, рассчитанная по восьми повторяющимся точкам, будет равна 6,42%.

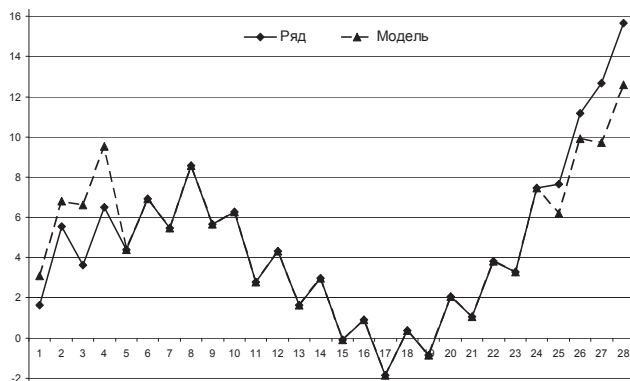


Рис. 4. Адекватность модели с линейным трендом

Строим окончательный прогноз. Для этого предлагается использовать процедуру экспоненциального сглаживания [4]:

$$F_t^{np} = \alpha F_{t-1}^{\phi} + (1-\alpha)F_t^m, \tag{8}$$

где F_t^{np} – прогноз; F_{t-1}^{ϕ} – фактическое значение в предыдущем аналогичном сезоне; F_t^m – значение модели; α – константа сглаживания.

Таким образом, прогноз на 29-й интервал определяется следующим образом.

Определяем прогнозное значение модели:

$$F_{29}^m = 0,1288 \cdot 29 + 2,1931 + 1,5392 \pm 6,42\% = 7,4674 \pm 0,4794 \tag{9}$$

Фактическое значение в предыдущем аналогичном сезоне составило

$$F_5^{\phi} = 4,3762 \tag{10}$$

Примем коэффициент сглаживания $\alpha = 0,2$. Получим прогнозное значение:

$$F_{29}^{np} = 0,2 \cdot F_5^{\phi} + (1-0,2)F_{29}^m = 0,2 \cdot 4,3762 + (1-0,2)(7,4674 \pm 0,4794) = 6,8492 \pm 0,3835. \tag{11}$$

Аналогичным образом определим прогнозы на 23 будущих периода (рис. 5). Качество прогноза неудо-

влетворительное, что выражается, во-первых, в высоком темпе нарастания ошибки, а во-вторых, в наличии значительного разрыва исходного ряда и первого прогнозного значения (29-я точка).

Аналогичным методом построим модель с экспоненциальным трендом. Его уравнение примет вид:

$$T = 4,8761 \cdot e^{0,0028t} - 1,0185. \tag{12}$$

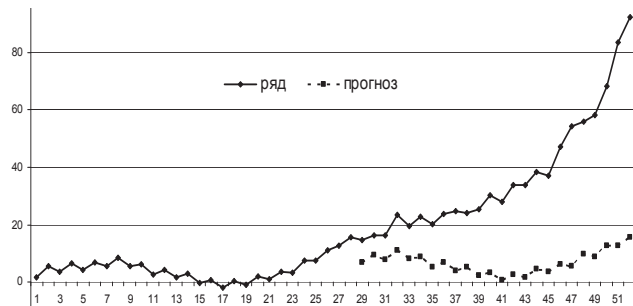


Рис. 5. Прогноз на сезон на основе модели с линейным трендом

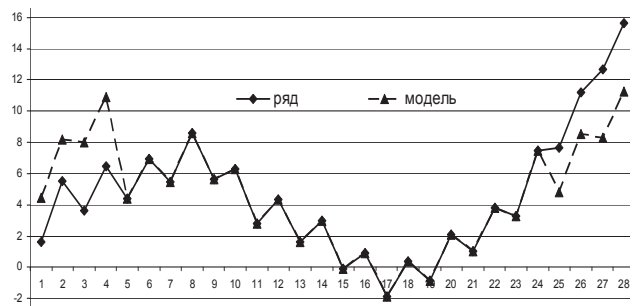


Рис. 6. Адекватность модели с экспоненциальным трендом

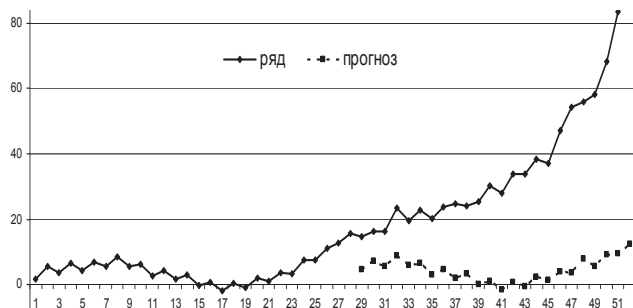


Рис. 7. Прогноз на сезон на основе модели с экспоненциальным трендом

Как видно из рис. 7, прогноз на основе модели с экспоненциальным трендом, найденным с помощью MS Excel, еще менее точен, чем с линейным.

Таким образом, можно убедиться, что при отсутствии информации о точном значении периода сезонности и малой длине ряда, модель прогнозирования Тейла-Вейджа обладает низкой точностью и становится практически неприменимой.

Воспользуемся теперь программным продуктом CaterpillarSSA 3.40, реализующим метод «Гусеница-SSA». В ходе анализа исходного ряда получено указанное на рис. 8 расположение корней характеристического полинома линейной рекуррентной формулы.

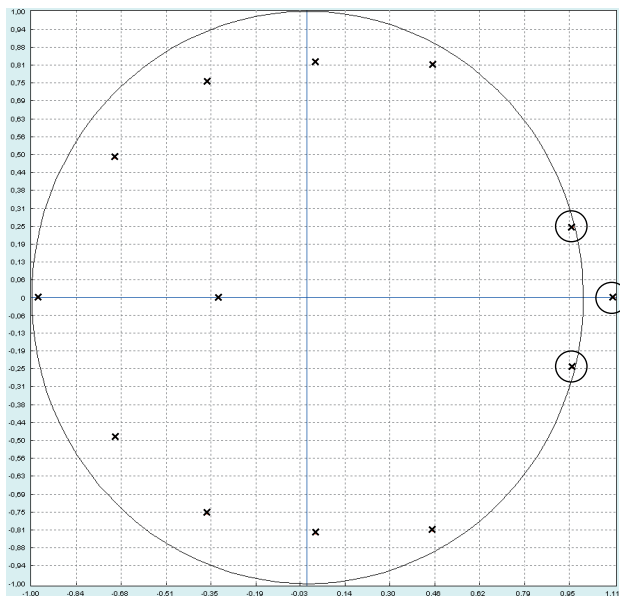


Рис. 8. Расположение корней характеристического полинома

Руководствуясь основными рекомендациями метода «Гусеница-SSA», предположим, что три корня, находящиеся в правой части рис. 8 (обведены кружками), можно отнести к группе так называемых главных корней, т.е. характеризующих реальную динамику процесса. Причем корень на вещественной оси соответствует медленно меняющейся составляющей ряда, а пара комплексных корней, ближайших к точке (1;0) и лежащих на единичной окружности – гармонической составляющей с наименьшей частотой [5]. Значение частоты этой гармонике используем для определения периода сезонности:

$$\operatorname{tg} \omega = \frac{\operatorname{Im}}{\operatorname{Re}}, \quad (13)$$

$$l = \frac{2\pi}{\omega} = \frac{2\pi}{\operatorname{arctg} \frac{\operatorname{Im}}{\operatorname{Re}}} = \frac{2\pi}{\operatorname{arctg} \frac{0,24606}{0,96786}} = 25,2379. \quad (14)$$

Для определения параметров тренда используем координаты вещественного корня:

$$\beta = \ln(\operatorname{Re}) = \ln(1,10076) = 0,096, \quad (15)$$

$$T = Ce^{\beta t} = e^{0,096t}. \quad (16)$$

Воспользуемся полученными частотными оценками и повторим описанную процедуру построения аддитивной модели (рис. 9, 10).

$$l = 25, T = e^{0,096t} - 0,92998. \quad (17)$$

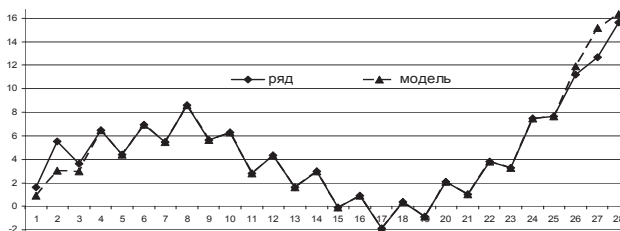


Рис. 9. Адекватность модели с частотными параметрами, полученными методом «Гусеница-SSA»

При этом среднеквадратическая ошибка, рассчитанная по шести повторяющимся точкам, составила 2,21%.

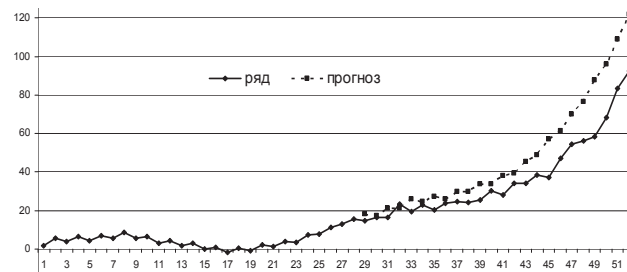


Рис. 10. Прогноз на сезон на основе смешанной модели

Ошибка прогноза существенно уменьшилась, разрыв между первым прогнозом и значением ряда отсутствует, что говорит об удовлетворительной степени адекватности предложенного подхода. Учитывая, что значение внутреннего параметра экспоненциального сглаживания α было выбрано неоптимальным, возможно добиться большей точности прогноза, воспользовавшись поисковыми настройками процедурами.

5. Выводы

Таким образом, в результате решения задачи краткосрочного прогнозирования процесса с сезонной динамикой были получены следующие результаты:

- 1) длина ряда менее двух полных периодов сезонности служит препятствием для корректного использования любого из рассматриваемых методов прогнозирования;
- 2) метод «Гусеница-SSA» позволяет определять значение периода сезонности с высокой точностью даже в зашумленных рядах (в модельном примере максимальный уровень шума – 10%, погрешность определения частоты основной гармонике – менее 1%);
- 3) кроме периода сезонности, целесообразно также определять методом «Гусеница-SSA» параметры тренда;
- 4) точность смешанного подхода, основанного на идентификации частотных характеристик и передаче их из модели «Гусеница-SSA» в модель Тейла-Вейджа, выше точности этих методов, используемых по отдельности.

Литература

1. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: Учеб. пособие. - М.: Финансы и статистика, 2003.-416 с.
2. Ханк Д.Э., Уичерн Д.У., Райтс А.Дж. Бизнес-прогнозирование, 7-е издание. : Пер. с англ. – М.: Издательский дом "Вильямс", 2003. – 656 с.
3. Кошечкин С.А. Алгоритм прогнозирования объема продаж в MS Excel – http://www.cfin.ru/finanalysis/sales_forecast.shtml.

4. Вартанян В.М., Кашеева В.Ю. Обзор методов прогнозирования показателей бизнес-процессов по временным рядам / Экономика та управління підприємствами машинобудівної галузі: проблеми теорії та практики – Харьков: Нац. аэрокосм. ун-т «ХАИ». – 2010. Додаток. Тези доповідей VIII Міжнародної науково-практичної конференції „Сучасні інформаційні технології в економіці та управлінні підприємствами, програмами та проектами. С. 189-191.
5. Голяндина Н.Э. Метод "Гусеница", - SSA: анализ временных рядов: СПб.: С. Петербургский гос. Университет, 2004. – 76 с.

Запропоновано підхід до розвитку інформаційних систем, що базується на розробці, яка спонукувана поведженням. Розглянуті фреймворки, орієнтовані на дану методологію

Ключові слова: IT-проект, розвиток інформаційних проектів, що орієнтовані на поведження, гнучкі методології розвитку

Предложен подход к развитию информационных систем, базирующийся на разработке, движимой поведением. Рассмотрены фреймворки, ориентированные на данную методологию

Ключевые слова: IT-проект, развитие информационных проектов, ориентированное на поведение, гибкие методологии развития

The approach for information systems' evolution which based on Behavior Driven Development is proposed. Frameworks that oriented on this methodology are considered

Key words: IT-project, Behavior Driven Development of IT-projects, Agile development

УДК 338.27.015

ФРЕЙМВОРКИ ДЛЯ ПРОЕКТИРОВАНИЯ ИНФОРМАЦИОННЫХ ПРОЕКТОВ, БАЗИРУЮЩИЕСЯ НА МЕТОДОЛОГИИ BDD

Н. В. Шатохина

Кандидат технических наук, доцент

Контактный тел.: 050-230-38-67

E-mail: shatosha@mail.ru

О. Я. Никонов

Доктор технических наук, заведующий кафедрой

Кафедра информатики

Харьковский национальный автомобильно-дорожный университет

ул. Петровского, 25, г. Харьков, Украина, 61002

Контактный тел.: (057) 707-36-58

Е. И. Антоненкова*

*Кафедра стратегического управления

Национальный технический университет "Харьковский политехнический институт"

ул. Фрунзе, 21, г. Харьков, Украина, 61002

Контактный тел.: (057) 707-68-24

1. Постановка проблемы

Современная индустрия разработки программных продуктов достигла высокого уровня зрелости. Тем не менее, процент успешных информационных проектов невелик. Менеджеры IT-проектов сталкиваются с такими проблемами как: ролевые конфликты (например, между бизнес-аналитиками и разработчиками); коммуникационные проблемы; снижение показателей эффективности труда; несогласованность нефункциональных требований к системе, а также отсутствие регламентированной процедуры их изменения. Оче-

видно, что причина всех этих проблем кроется в отсутствии методологически обоснованного планирования и управления проектом.

Процесс управления требованиями традиционно считается одним из ключевых при создании информационных проектов – наибольшие риски проектов связаны с высокой изменчивостью требований и ошибками в их определении.

Для обеспечения качественного развития информационных проектов в условиях изменчивого внешнего окружения, которое проявляется присутствием рисков и неопределенностью в пожеланиях заказчика,