6. Jain, A. K. Flynn Data clustering: a review [Text] / A. K. Jain, M. N. Murty // ACM Comput. Surv. – 1999. – Vol. 31(3). – P. 264–323.
7. Пістунов, І. М. Кластерний аналіз в економіці [Текст] / І. М. Пістунов, О. П. Антонюк та ін. – Дніпропетровськ: Національний гірничий університет, 2008.– 84 с.
8. Ким, Дж. Факторный, дискриминантный и кластерний анализ [Текст] / Дж. Ким, Ч. У. Мьюллер, У. Р. Клекка. – М.: Финансы и статистика, 1989. – 215 с.
9. Дюран, Б. Кластерный анализ [Текст] / Б. Дюран, П. Оделл. – М.: «Статистика», 1977. – 128 с.
10. Кондрук, Н. Е. Застосування багатокритеріальних моделей для задач збалансованого харчування [Текст] / Н. Е. Кондрук, М. М. Маляр // Вісник Черкаського державного технологічного університету. Серія: технічні науки. – 2010. – Вип. 1, № 1. – С. 3–7.
11. Кондрук, Н. Э. Некоторые применения кластеризации критериального пространства для задач выбора [Текст] / Н. Э. Кондрук, Н. Н. Маляр // Компьютерная математика. – 2009. – № 2. – С. 142–149.
12. A61K8/19, A61K8/30, МПК (2006.01). Патент на корисну модель 64777 Україна. Спосіб автоматизованого складання дієтичного харчування «Дієтолог» [Текст] / Маляр М. М., Кондрук Н. Е., Горленко О. М., Томей А.І . – № u201100007; Заявл. від 04.01.2011; Опубл. 25.11.2011, Бюл.№ 22.

УДК 517.534

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ПРОГНОЗИРОВАНИЯ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА

А. А. Чистякова
Аспирант*
E-mail: anna.chistyakova.prn@gmail.com
Б. В. Шамша
Кандидат технических наук, профессор*
E-mail: shamsha.b.v@gmail.com
*Кафедра информационных управляющих систем
Харьковский национальный университет радиоэлектроники
пр. Ленина, 16, г. Харьков, Украина, 61166

В роботі запропонована інформаційна технологія прогнозування нестаціонарних часових рядів, яка не зводяться до стаціонарних, характерізуються нелінійним трендом та завуальованимі періодичними компонентами. З метою побудови моделі прогнозування визначається поведінка компонент часового ряду у декілької фазових просторах, побудованних з використанням методу сингулярного спектрального аналізу (SSA)

Ключові слова: часовий ряд, прогнозування, інформаційна технологія, сингулярний спектральний аналіз, фазовий простір

В работе предложена информационная технология прогнозирования нестационарных временных рядов, которые не приводятся к стационарным, характеризуются нелинейным трендом и завуалированными периодическими компонентами. Для построения модели прогнозирования определяется поведение компонент временного ряда в нескольких фазовых пространствах, построенных с использованием метода сингулярного спектрального анализа (SSA)

Ключевые слова: временной ряд, прогнозирование, информационная технология, сингулярный спектральный анализ, фазовое пространство

1. Введение

Прогнозирование является одним из решающих элементов эффективной организации управления предприятиями вследствие того, что результат принимаемых решений в большой степени определяется качеством прогнозирования их последствий. Поэтому решения, принимаемые сегодня, должны опираться на достоверные оценки возможного развития изучаемых явлений, изменения технико-экономических показателей и событий в будущем.

Применение прогнозирования в информационных технологиях (ИТ) позволит воздействовать на ускоренный процесс анализа, обработки, распространения и использования обширной базы информации, а также своевременно принимать управленческие решения и планировать деятельность в соответствии с полученными результатами. На сегодняшний день информационные технологии выступают своего рода катализатором эволюции в мире. Поэтому применение ИТ в прогнозирования приводит к развитию научно-технического прогресса и к использованию результатов в хозяйственной практике.

В этой связи применение современных компьютерных технологий и разработка новых методов с их использованием обусловлена рядом причин, в числе которых:

– рост объемов информации;
– сложность структуры данных;
– появление современных методов прогнозирования, их развитие и внедрение;
– сложность алгоритмов расчета и интерпретации результатов;
– высокие требования к качеству прогнозов;
– необходимость использования результатов прогнозирования для решения задач планирования и управления.

Множество рядов технико-экономических показателей, отражающих рост предприятий и компаний, имеют нелинейный характер, что затрудняет их анализ, прогнозирование и принятие управленческих решений на основании статистических данных. Задача прогнозирования таких рядов становится все более и более актуальной на сегодняшний день, так как активно развиваются технологии, изменяются бизнес-процессы, существует большое количество внешних нелинейных аттракторов и т. д.

2. Постановка задачи

На сегодняшний день существует множество методов и подходов к прогнозированию временных рядов, среди которых можно выделить наиболее современные и широко использующиеся в практических задачах. К ним относят: метод ARIMA, классификационно-регрессионные алгоритмы, метод максимального правдоподобия, метод идентификации кластеров на основании статистических характеристик, метод GARCH, нейросетевой метод и др. Каждый из данных методов имеет свои предпосылки и предположения при практическом использовании. Так, в частности, для построения авторегрессионных моделей необходимо, чтобы временной ряд был стационарным или приводился к таковому путем взятия разности первого, второго или большего порядков. Метод GARCH применяется для и построения моделей волатильности временного ряда, нейросетевые методы требуют достаточного, но не избыточного количества данных обучающей последовательности, а также выбора адаптивной структуры нейросети. Другие методы имеют свои особенности и требуют своих предпосылок и предположений. Однако, нестационарные временные ряды характеризуются нелинейностью, сезонной компонентой со смещением, периодическими компонентами с изменяющейся амплитудой, частотой, которые слабо разделимы с компонентой шума, что затрудняет их прогнозирование и оценку доверительного интервала.

В виду выделенных особенностей нестационарные временные ряды стоит рассматривать не только как набор некоторых детерминированных компонент, но и как реализацию данных компонент в различных состояниях системы. В работе предлагается разработать информационную технологию для решения задачи прогнозирования нестационарных временных рядов, которые не приводятся к стационарным и характеризуются неоднородной структурой, с использованием метода сингулярного спектрального анализа. Целью данного исследования является построение модели прогнозирования сложных стохастических временных рядов с учетом нескольких фазовых пространств при отсутствии априорной информации о структуре ряда.

Реализация предложенного метода, построение модели временного ряда и оценка результатов выполнена с использованием языка программирования R в среде R-studio.

3. Анализ литературных данных

Традиционные подходы к проблеме прогнозирования достаточно подробно изложены в научных трудах Айвазяна С. А., Андерсена Т., Бендита Дж., Бокса Дж., Беллинджера Д., Дженкинса Г., Кенделла М., Монтгомери Б., Пирсола А., Степанова В. С., Стюарта А., Уолкера Г. [1 – 3]. Наиболее известными пакетами прикладных программ, реализующими традиционные методы, модели и алгоритмы прогнозирования, являются: CSS, Deductor, Forecast Expert, Predictor, SAS, S-plus, SPSS, STATISTICA, STATGRAPHICS. Однако, эти пакеты не могут быть использованы для прогнозирования нестационарных временных рядов без предварительной обработки и подготовки данных, зачастую используют ансамбль методов для построения модели прогнозирования.

В работе рассматривается метод сингулярного спектрального анализа (SSA), который появился в 70-90-г.г. XX века и привлек внимание многих исследователей, как новая и эффективная техника анализа структуры временных рядов [4, 5]. Метод включает элементы классического анализа временных рядов, многомерного статистического анализа, многомерной геометрии, методы динамических систем и обработки сигналов. Области применения SSA: от математики и естественных наук до экономики и банковской сферы, от метеорологии и океанологии до социологии и исследования рынка. Основные этапы метода реализованы с помощью языка программирования R [6].

Метод прогнозирования с использованием сингулярного спектрального анализа в последние несколько лет привлек внимание множества исследователей из различных предметных областей [7 – 10]. Однако, в работах рассматриваются проблемы прогнозирования стационарных временных рядов или таковых, которые приводятся к стационарным. Таким образом вопрос построения модели прогнозирования нестационарных временных рядов является нерешенным.

4. Описание метода SSA

Отличительной особенностью и достоинством метода SSA является отсутствие требования априорного задания модели временного ряда, метод является адаптивным и не требует стационарности исходных данных. К преимуществам метода также можно отнести возможность выделения нелинейного тренда и работу с модулированными гармониками.

Основными шагами метода SSA являются:

– построение траекторной матрицы X с выбором глубины погружения или длины окна L, который также называют этапом вложения;

– сингулярное разложение матрицы XX^T

$$X = X_1 + X_2 + \cdots + X_d, X_i = \sqrt{\lambda_i} U_i V_i, \quad (1)$$

где $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d > 0$ – упорядоченные ненулевые собственные числа матрицы XX^T, $\{U_i\}_{i=1}^d : U_i \in \Re^L$;

– выбор главных компонент на основании собственных векторов матрицы и их группировка;

– восстановление временного ряда путем усреднения матрицы $\tilde{X}$, построенной на основании выбранных главных компонент, по ее побочным диагоналям.

Подробно метод изложен в работе [11].

Входными параметрами метода являются временной ряд Y и глубина погружения L. Для определения фазового пространства временного ряда необходимо определить его размерность, которая определяется параметром L, представив его в виде пространства ортонормированных векторов. Вопрос определения размерности, при которой фазовое пространство наилучшим образом демонстрирует характеристики нестационарного временного ряда подробно рассматривается в [12].

Самым неформализуемым шагом метода SSA является шаг группировки. Вся информация о каждой из компонент X_i содержится в собственном числе λ_i а также в собственном U_i и факторном V_i векторах, которые имеют определенную структуру в построенном фазовом пространстве временного ряда Y. Собственный и факторный вектора называют сингулярными векторами, а совокупность $(\sqrt{\lambda_i}, U_i, V_i)$ – собственной тройкой, каждая из которых отражает величину дисперсии λ_i в направлении U_i. Поиск компонент для требуемой группировки, главным образом на основе анализа собственных троек, будем называют процедурой идентификации.

Для идентификации компоненты тренда нестационарного временного ряда в работе используется анализ автокорелляционных функций собственных векторов. Если характер автокорреляционной функции является монотонно убывающим, то гипотеза о компоненте тренда принимается, иначе отвергается.

Для идентификации периодических составляющих используется периодограммный метод, суть которого состоит в оценке спектральной плотности с применением преобразования Фурье. Проводится анализ пиков периодограмм собственных векторов, если два соседних собственных вектора имеют единственный пик на периодограмме, который соответствует одинаковой частоте, то гипотеза о периодической компоненте принимается, иначе отвергается.

В результате построения фазового пространства одномерного временного ряда Y и выделения детерминированных компонент на этапе восстановления получают модель временного ряда и его составляющих, что является важным этапом при прогнозировании нестационарных временных рядов. Метод SSA является робастным методом и может быть применен в качестве фильтра с настраиваемой пропускной способностью.

4. 1. Информационная технология прогнозирования

Прогнозирование временного ряда F возможно, если ряд имеет определенную структуру, выбран метод нахождения структуры временного ряда, с помощью которого ряд может быть продолжен, и спрогнозированы его значения, структура модели ряда сохраняется в период прогнозирования. Определить и обеспечить стабильность модели для прогнозируемых значений практически невозможно, так как возникают новые флуктуации, влияющие на поведение временного ряда, а также непредсказуемые интервенции. Наличие шумовой компоненты во временном ряду также влияет на прогнозируемые значения.

Для нестационарных временных рядов, которые характеризуются нелинейной динамикой и не приводятся к стационарным, применение классических методов прогнозирования является некорректным, так как нарушаются предпоссылки и предположения методов прогнозирования. Метод SSA не требует задания структуры модели ряда, в следствии этого особый интерес представляет использование данного метода для прогнозирования нестационарных временных рядов.

Рассмотрим наиболее распространённые структуры моделей нестационарных временных рядов. В модели "сигнал плюс шум" элемент представляется следующим образом:

$$f_n = s_n + \varepsilon_n, \quad s_n = S(s_T, \ldots, s_{n-1}), \quad (2)$$

то есть можно представить такой ряд, как детерминированный сигнал $s_T, \ldots, s_{n-1}$, наблюдения которого проводятся с ошибкой ε_n. Сигнал в общем случае может быть представлен с помощью рекуррентной формулы S.

Для процессов авторегрессии (порядка p) полагают, что функция F имеет линейный вид и f_n зависит от p предыдущих значений, случайная шумовая составляющая имеет нулевое среднее и постоянную ковариацию. Таким образом, параметрами такого процесса, кроме ковариации, будут коэффициенты $a_1, \ldots, a_p$ функции F:

$$f_n = a_1 f_{n-1} + \ldots + a_p f_{n-p} + \varepsilon_n. \quad (3)$$

Подход, подобный авторегрессионному, используется и в моделях "сигнал плюс шум" следующим образом: полагают, что сигнал ряда задается линейной рекуррентной формулой (ЛРФ):

$$f_n = s_n + \varepsilon_n, \quad s_n = a_1 s_{n-1} + \ldots + a_p s_{n-p}. \quad (4)$$

Такая модель тоже имеет p неизвестных параметров $a_1, \ldots, a_p$ вдобавок к параметрам распределения шума. Наложим следующие стандартные ограничения на шум: $\varepsilon_0, \ldots, \varepsilon_{n-1}$ независимы, одинаково распределены, $E(\varepsilon_n) = 0$ и $D(\varepsilon_n) = \sigma^2$. Для определения коэффициентов в рамках метода SSA необходимо выбрать пространство определенной размерности (эквивалентно выбору глубины погружения L) и идентифицировать главные детерминированные компоненты временного ряда в данном пространстве.

При построении модели прогнозирования методом SSA предполагается, что временной ряд описывается с помощью линейных рекуррентных формул (ЛРФ), то есть является линейной комбинацией экспонент, полиномов и гармоник. Класс таких рядов достаточно широк. Доказано [13], что траекторное пространство порождает ЛРФ порядка L-1, если размерность данного пространства $L \times K$. При нахождении фазового пространства временного ряда с определенной глубиной погружения L можно построить ЛРФ управляющую рядом в пределах данного пространства. Покажем, как с помощью главных собственных векторов можно найти авторегрессионные коэффициенты порядка p, используя известные результаты метода SSA о продолжении восстановленных компонент [13]. В методе SSA для определения коэффициентов модели прогнозирования используется формула:

$$a = \frac{1}{1 - v^2} \sum_{j=1}^{r} u_{L,j} u_j^{L-1} . \qquad (5)$$

где $v = u_{L,1}^2 + u_{L,2}^2 + \ldots + u_{L,r}^2$ является суммой квадратов последних элементов собственных векторов $j = 1, 2, \ldots, r$ (вертикальные коэффициенты), а обозначает $(k - 1 \times 1)$ вектор с первыми L-1 элементами собственных векторов $j = 1, 2, \ldots, r$. Продолжение восстановленной компоненты ряда возможно если $v^2 < 1$.

Представим обобщенную информационную технологию прогнозирования.

Входные данные:

– временной ряд F_N и/или его модель $\tilde{F}_N$;

– сингулярное разложение траекторной матрицы временного ряда F_N, построенное с фиксированной глубиной погружения L: $S = \sqrt{\lambda} U V^T$;

– выбранные номера компонент $(g \in G)$, которые соответствуют детерминированному сигналу тренда или периодической компоненты и образуют базис главных компонент временного ряда;

– количество точек прогноза M.

Шаг 1. Построение линейной рекуррентной формулы (вектора) ЛРФ.

1. 1. Построение матрицы U_G из собственных векторов детерминированного сигнала размером $L \times r$ (где r – размерность множества G)

1. 2. Нахождение вектора $lpf = U_G \times U_{G_L}$, где U_{G_L} - последняя строка матрицы U_G.

1.3. Получение коэффициентов ЛРФ $lrr = (a_1, ..., a_{L-1})$ путем масштабирования (преобразования) вектора lpf:

1. 4.

$$lrr = lpf^T / (1 - lpf_n) , \qquad (6)$$

где lpf_n - последний элемент вектора

Шаг 2. Построение M точек прогноза, последовательно применяя одну из формул:

$$F_{N+1} = \sum_{i=1}^{L-1} F_{N+i-L} \times lrr_i , \qquad (7)$$

$$\tilde{F}_{N+1} = \sum_{i=1}^{L-1} \tilde{F}_{N+i-L} \times lrr_i . \qquad (8)$$

В первом варианте (7) для построения прогноза используется исходный временной ряд, во втором (8) – модель временного ряда.

В результате выполнения всех шагов предложенной информационной технологии строится линейная авторегрессионная модель прогнозирования, коэффициенты которой учитывают особенности нестационарного временного ряда F, а также L-1 последних лагов исходного временного ряда или построенной ее модели с использованием метода SSA.

4. 2. Информационная технология построения фазовых портретов нестационарных временных рядов

Возникает несколько проблем прогнозирования временного ряда с использованием ЛРФ:

– стабильность и состоятельность выбранного фазового пространства;

– корреляция коэффициентов ЛРФ и значений временного ряда, что зачастую приводит к значительной ошибке прогноза;

– оценка точности прогнозируемых данных.

Для решения первой проблемы предлагается провести анализ различных фазовых пространств временного ряда F с использованием разной глубины погружения. При относительно небольшой величине параметра L более точно проводится идентификация тренда, при $L \to N/2$ высокочастотные и низкочастотные части спектра хорошо разделяются. При выборе параметра глубины погружения с учетом сезонной составляющей необходимо выбрать величину L кратную периоду определенной сезонности. Для построения модели с высоким коэффициентом детерминации и определения недетерминированной случайной компоненты данный параметр выбирается минимальным из допустимых, т. е. $L \to 1$. Последний подход используется при идентификации модели «сигнал+шум».

При построении нескольких фазовых пространств одномерного нестационарного временного ряда необходимо определить оптимальное их количество и размерности. Данная задача решается методом анализа суммарных коэффициентов корреляции независимых векторов траекторных матриц. Независимыми векторами назовем те, которые не пересекаются и взяты последовательно. Опишем информационную технологию выбора параметров глубины погружения в методе SSA для прогнозирования нестационарных временных рядов на основании нескольких фазовых пространств или портретов. Данная информационная технология реализована с помощью языка программирования R, что позволяет использовать различные хранилища данных для импорта информации.

Шаг 1. Импорт данных из базы данных или в формате .csv, .xls, .dat, .sav, .txt.

Шаг 2. Предварительный анализ данных, восстановление пропущенных значений, подготовка вектора временного ряда для дальнейшего анализа.

Шаг 3. Построение множества $N/2$ траекторных матриц с использованием параметра L, $L \in [1; N/2]$, где N – длина временного ряда.

Шаг 4. Вычисление среднего коэффициента корреляции независимых векторов траекторной матрицы $E(r_L)$, как математического ожидания степени линейной зависимости между соседними участками временного ряда одинаковой длины L.

Шаг 5. Построение вектора $V(E(r_L))$, из множества $E(r_L)$, каждый элемент которого строго соответствует определенному пространству временного ряда при $L \in [1; N/2]$.

Шаг 6. Определение индексов локальных минимумов вектора $V(E(r_L))$, которые достигнуты при $I = ind_{min(V(E(r_L)))}$.

Шаг 7. Построение нескольких фазовых пространств временного ряда при $L = I$ с использованием главных детерминированных компонент данных пространств, полученных с помощью информационной технологии идентификации структуры временно ряда методом SSA.

В работе предлагается использовать K фазовых пространств, соответствующие $L \in \{L_1, ..., L_K\}$ и построенные ЛРФ, как продолжения временного ряда в разных базисах. Таким образом, адекватность модели прогнозирования может быть улучшена, так как учитывается несколько состояний системы в разных K базисах, порождаемых самим временным рядом. Данный подход повысит устойчивость прогноза и определит сбалансированное влияние ортонормированных векторов при различных состояниях системы, а также позволит учитывать разное количество последних лагов в соответствии с глубиной погружения, размерностью фазового пространства и главными компонентами временного ряда.

4. 3. Информационная технология прогнозирования нестационарных временных рядов с использованием нескольких фазовых пространств

В работе предлагается модифицировать модель прогнозирования временного ряда, используя выбранные K фазовых пространств временного ряда. При выборе параметра глубины погружения и построении матриц разных размерностей может наблюдаться приблизительно одинаковая слабая разделимость компонент ряда, что показывает средний коэффициент корреляции траекторных матриц. Для построения модели прогнозирования нестационарных временных рядов необходимо рассчитать K наборов коэффициентов ЛРФ, отражающих поведение главных компонент ряда в каждом выбранном фазовом пространстве, и задать весовые коэффициенты оценки разделимости в данном фазовом пространстве $w_k, k \in [1; K]$. Если ряд имеет приблизительную равную слабую разделимость во всех базисах, то коэффициенты могут быть равны, однако, необходимо выполнение следующих условий: $w_k < 1$ и $\sum_{k=1}^{K} w_k = 1$. Тогда для построения 1-й точки прогноза временного ряда F_N используем модифицированную ЛРФ:

$$F_{N+1} = \sum_{i=1}^{K} w_i R_i, \qquad (9)$$

$$R_i = \sum_{j=1}^{P_i} F_{N+j-P_i} \times lrr_{i_j}, \qquad (10)$$

где P_i – соответствует размерности i-го фазового пространства, фактически является количеством коэффициентов ЛРФ в данном i-м пространстве;

lrr_{i_j} – коэффициенты ЛРФ в i-м фазовом пространстве;

w_i – весовой коэффициент i-го фазового пространства.

Используя модифицированную формулу прогнозирования временного ряда можно получить несмещенную и устойчивую оценку прогнозируемого временного ряда.

Коэффициенты ЛРФ в i-м фазовом пространстве определяются по методу изложенному выше при описании информационной технологии прогнозирования в выбранном пространстве нестационарного временного ряда с определенным параметром L.

5. Экспериментальные данные и их обработка

В работе покажем реализацию информационной технологии идентификации структуры и прогнозирования нестационарного временного ряда в различных фазовых пространствах. Построение модели прогнозирования с использованием метода SSA и предложенной информационной технологии реализовано посредством языка программирования R-programming в среде R-studio.

В качестве входных данных используем временной ряд курсов валют EUR/USD с 2012-12-1 по 2013-12-9, обозначим его F_N, график данных представлен на рис. 1.

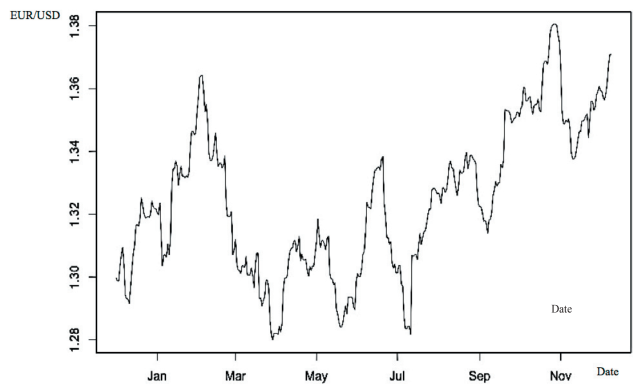

Рис. 1. График временного ряда курсов валют EUR/USD

Для проверки ряда на стационарность применим тест KPSS к исходным данным, а также разностям первого и второго порядков, который является расширенным тестом Дикки-Фуллера. Результаты представлены в табл. 1.

Таблица 1

Результаты теста KPSS к исходным данным, а также разностям первого и второго порядков

	Результат теста KPSS
F_N	KPSS Level = 2.8924, p-value = 0.01
Первая разность	KPSS Level = 0.0673, p-value = 0.1
Вторая разность	KPSS Level = 0.0718, p-value = 0.1

Результаты теста показывают, что ряд не является стационарным и не может быть приведен к таковому.

Были определены параметры фазовых пространств L= [36, 75, 91, 121, 181], в которых данный временной ряд проявляет детерминированные свойства различного характера на разных его участках. В каждом фазовом пространстве ряд может быть кусочно аппроксимирован, а выделенные компоненты ряда показывают наиболее стабильные структуры данных, которые сохраняются в фазовых пространствах различной размерности.

Восстановим модели временного ряда с использованием разного параметра глубины погружения L. График восстановленных моделей ряда F_N представлен на рис. 2.

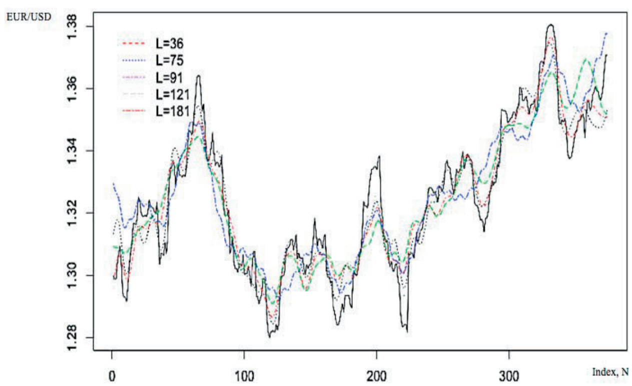

Рис. 2. График моделей временного ряда F_N, построенных с различным параметром глубины погружения L

Построенные модели описывают временной ряд, однако, с различной степенью «сглаживания». Безусловно, наблюдаются общие компоненты, например, компонента возрастающего тренда и некоторые выделенные периодики.

Так как временной ряд является нестационарным, что говорит о высокой степени изменчивости различного характера, можно утверждать, что все компоненты при различной величине погружения могут повлиять на продолжение ряда. В модели прогнозирования предлагается учитывать все выделенные на этапе сингулярного разложения главные компоненты в различных пространствах. Базисы всех выделенных фазовых пространств, которыми являются собственные вектора каждого из сингулярных разложений, являются основой аддитивной модели прогнозирования. Построенные фазовые пространства будут расширены, что представит продолжение нестационарного временного ряда в его различных состояниях.

Таким образом, влияние тех главные детерминированных компонент, которые дублируются в нескольких фазовых пространствах, будет усилено, а компоненты, выделенные ошибочно, будут элиминированы.

Рассмотрим алгоритм построения модели прогнозирования с технической точки зрения, используем информационную технологию построения прогноза временного ряда на основе ЛРФ и метода SSA в определенном базисе (траекторном пространстве) ряда F и модель прогнозирования, согласно формулам 5, 6, 9, 10, рассмотренные ранее.

Остановимся подробнее на методе определения весовых коэффициентов каждого из выбранных фазовых пространств в модели прогнозирования. Так как сумма весовых коэффициентов построенных ЛРФ должна равняться 1, то коэффициент каждого из фазовых пространств вычисляется на основании доли учтенной дисперсии в данном пространстве. Обозначим gr_i – индексы детерминированных собственных векторов в -м фазовом пространстве. Тогда:

$$w_i = \frac{\sum \lambda^i_{gr_i}}{\sum \lambda^i} \Big/ \sum_{j=1}^{K} \frac{\sum \lambda^j_{gr_j}}{\sum \lambda^j}, \qquad (11)$$

где λ^i – значения собственных чисел i -го сингулярного разложения, что соответствует i -му собственному пространству;

$\lambda^i_{gr_i}$ – значения собственных чисел соответствующих главным детерминированным собственным векторам i -го сингулярного разложения;

K – количество учитываемых фазовых пространств,

Построим модель прогнозирования с использованием 5-ти выделенных фазовых пространств, ЛРФ и SSA (формулы [9, 10]) и полученные коэффициенты (12). Длина исследуемого временного ряда равна N=374.

$$\begin{aligned} F_{N+1} &= 0.201426\sum_{j=1}^{36} F_{374+j-36}\times lrr_{1_j} + \\ &+0.200675\sum_{j=1}^{75} F_{374+j-75}\times lrr_{2_j} + \\ &+0.1993201\sum_{j=1}^{91} F_{374+j-91}\times lrr_{3_j} + \\ &+0.1994309\sum_{j=1}^{121} F_{374+j-121}\times lrr_{4_j} + \\ &+0.199148\sum_{j=1}^{181} F_{374+j-181}\times lrr_{5_j}. \end{aligned} \qquad (12)$$

Модель(12) описывает временной ряд F_N с помощью ЛРФ с использованием различной величины глубины погружения. Расчет прогноза проводится последовательно на 1 лаг.

График прогноза нестационарного временного ряда курсов валют на 7 лагов представлен на рис. 3

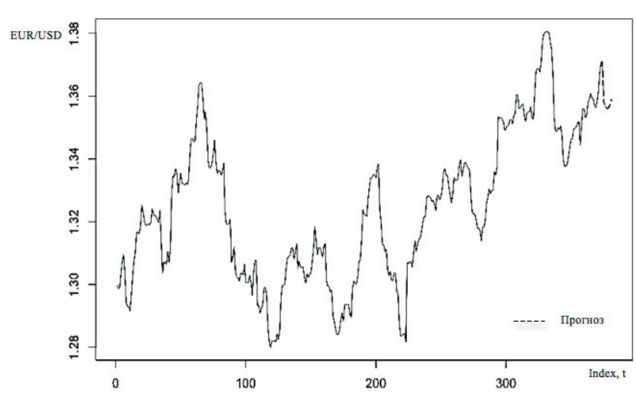

Рис. 3. Модель прогнозирования нестационарного временного ряда курсов валют на 7 лагов

Для построения модели прогнозирования нестационарного временного ряда курсов валют были использованы пять фазовых пространств, что позволило определить и учесть его наиболее стабильные компоненты.

6. Выводы

В работе предложена информационная технология прогнозирования нестационарных временных рядов, которые не могут быть приведены к стационарным. Идентификация структуры временного ряда проводится с и применением метода SSA, так как данный метод не требует априорной информации о характере ряда. Для построения модели прогнозирования предложено использовать нескольких фазовых пространств исходных данных, что позволяет учитывать поведение компонент временного ряда в различных состояниях системы. Построение фазовых пространств проводится путем представления одномерного временного ряда в новом ортогональном базисе, оси которого ориентированы по направлениям максимальной дисперсии набора входных данных. При изменении параметра глубины погружения в методе SSA временной ряд может быть представлен в различных базисах, в каждом из которых такие компоненты, как тренд и периодическая составляющая имеют разный характер.

Предполагается, что временной ряд описывается набором линейных рекуррентных формул, коэффициенты каждой из которых определены в некотором фазовом пространстве. При идентификации и использовании детерминированных компонент временного ряда, в нескольких базисах можно выделить наиболее перманентные и устойчивые, что влияет на качество прогноза. Построенная модель прогнозирования учитывает различную глубину погружения и весовые коэффициенты каждого из построенных фазовых портретов нестационарного временного ряда, что позволяет получить адекватные оценки прогноза.

Реализация предложенной информационной технологии проведена на языке программирования R и может быть использована с целью построения краткосрочного прогноза для нестационарных временных рядов.

Литература

1. Айвазян, С. А. Прикладная статистика и основы эконометрики [Текст] / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – 1006 с.
2. Box, G. E. P. Time Series Analysis: Forecasting and Control [Text] : 4rd ed / G. E. P. Box, G. M. Jenkins, G. C. Reinsel. – US: John Wiley & Sons, 2008.
3. Kendall, M. Design and analysis, and time series, the advanced theory of statistics [Text] ; 3rd ed./ M. Kendall, A. Stuart // London: Charles Griffin. – 1976. – Vol. 3.
4. Данилов, Д. Л. Главные компоненты временных рядов: Метод "Гусеница" [Текст] / Д. Л. Данилов, А. А. Жиглявский. – СПб.: Изд. дом "ПРЕССКОМ", 1997. — 307 с.
5. Vautard, R. Singular-Spectrum Analysis: a toolkit for short, noisy chaotic signals [Text] / R. Vautard, P. Yiou, M. Ghil // Physica D. – 1992. – №58. – P. 95–126.
6. Golyandina, N. Basic Singular Spectrum Analysis and forecasting with R [Text] / N. Golyandina, A. Korobejnikov // Computational Statistics & Data Analysis. – 2014. – № 71. – P. 934–954.
7. Nekrutkin, V. Perturbation expansions of signal subspaces for long signals [Text] / V. Nekrutkin // Statistics and Its Interface. – 2010. – №3. – P. 297–319.
8. Hassani, H. Singular Spectrum Analysis: Methodology and Application to Economics Data [Text] / H. Hassani, A. Zhigljavsky // Journal of System Science and Complexity. – 2009. – № 22(3). – P. 372–394.
9. Briceño, H. Singular Spectrum Analysis for Forecasting of Electric Load Demand [Text] / H. Briceño, C. M. Roccoa, E. Zio // Chemical endineering transactions. – 2013. – № 33. – P. 919–924.
10. Pepelyshev, A. Assessing the stability of long-horizon SSA forecasting [Text] / A. Pepelyshev, A. Zhiglyavsky // Statistics and Its Interface. – 2010. – №3. – P. 321–327.
11. Чистякова, А. А. Идентификация структуры нестационарного временного ряда при помощи метода сингулярного спектрального анализа [Текст] / А. А. Чистякова, Б. В. Шамша // Радиоэлектронные и компьютерные системы. – 2011. – № 4(52). – С. 105–111.
12. Чистякова, А. А. Оценка глубины погружения в методе SSA при моделировании нелинейных временных рядов [Текст] / А. А. Чистякова, Б. В. Шамша // Вестник развития науки и образования. – 2013. – № 4. – С. 59–68.
13. Голяндина, Н. Э. Метод «Гусеница»-SSA: прогноз временных рядов [Текст] : уч. пос. / Н. Э. Голяндина. – СПб., 2004. – 52 с.